

# Psychometrika

---

## CONTENTS

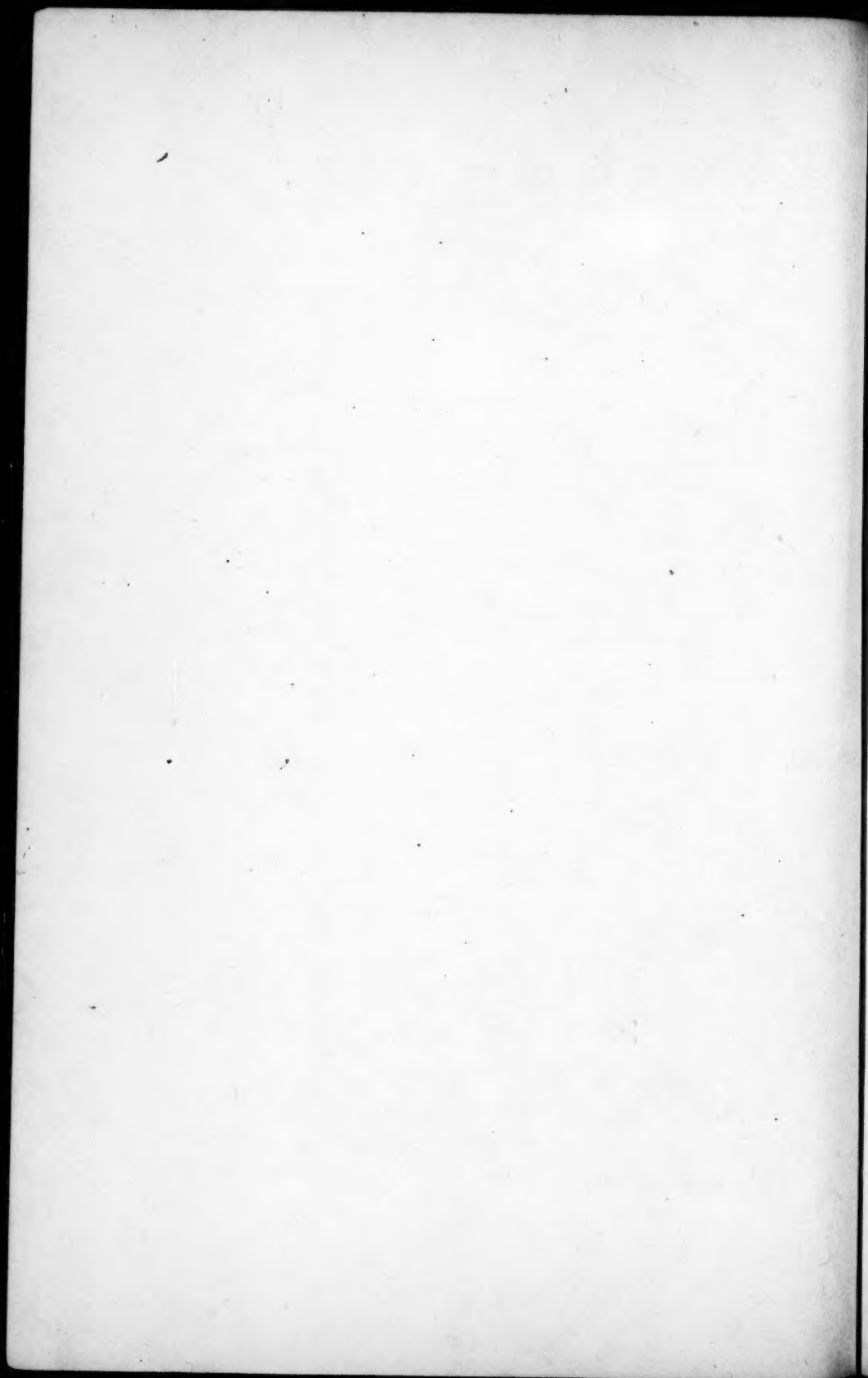
GOODNESS OF FIT OF TREND CURVES AND SIGNIFICANCE OF TREND DIFFERENCES - - - -	65
E. F. LINDQUIST	
THE ESTIMATION OF RELIABILITY WHEN SEVERAL TRIALS ARE AVAILABLE - - - - -	79
HOWARD W. ALEXANDER	
AN EMPIRICAL COMPARISON OF METHODS OF COMMUNALITY ESTIMATION - - - - -	101
FRANCIS F. MEDLAND	
TABLES FOR THE DETERMINATION OF THE SIGNIFICANCE OF SKEWNESS AND OF THE SIGNIFICANCE OF THE DIFFERENCE IN THE SKEWNESS OF TWO INDEPENDENT DISTRIBUTIONS -	111
ERWIN K. TAYLOR	
A MODIFIED AITKEN PIVOTAL CONDENSATION METHOD FOR PARTIAL REGRESSION AND MULTIPLE CORRELATION - - - - -	127
ALICE VAN BOVEN	
DYNAMICS OF THE CEREBRAL CORTEX: AUTOMATIC DEVELOPMENT OF EQUILIBRIUM IN SELF-ORGANIZING SYSTEMS - - - - -	135
W. R. ASHBY	
TABLE OF MEAN DEVIATES FOR VARIOUS PORTIONS OF THE UNIT NORMAL DISTRIBUTION - -	141
HOLLIS M. LEVERETT	
NOTE ON THE SAMPLING ERROR OF THE DIFFERENCE BETWEEN CORRELATED PROPORTIONS OR PERCENTAGES - - - - -	153
QUINN MCNEMAR	
REVIEWS - - - - -	159
MEMBERS OF THE PSYCHOMETRIC SOCIETY - - -	163

---

VOLUME TWELVE

JUNE 1947

NUMBER TWO



## GOODNESS OF FIT OF TREND CURVES AND SIGNIFICANCE OF TREND DIFFERENCES

E. F. LINDQUIST  
STATE UNIVERSITY OF IOWA

Experimental studies of the successive changes (frequently represented by curves describing laws of learning and other similar functional relationships) in a criterion variable accompanying experimental variations in a given "treatment," and experimental comparisons of such changes for different populations or for different treatments, constitute a large and important class of psychological experiments. In most such experiments, no attempt has been made to analyze or to make allowance for errors of sampling or of observation. In many others, the techniques of error analysis that have been employed have been inefficient, inexact, or inappropriate. This paper suggests tests, using the methods of analysis of variance, of certain hypotheses concerning trends and trend differences in sample means obtained in experiments of this general type. For means of successive independent samples, tests are provided of the hypotheses: ( $H_1$ ) that there is no trend, or that the trend is a horizontal straight line, ( $H_3$ ) that there is a linear trend, ( $H_5$ ) that the trend is as described by a line not derived from the observed means, and ( $H_7$ ) that the trend is as described by a line fitted to the observed means. Tests are also provided of similar hypotheses ( $H_2$ ,  $H_4$ ,  $H_6$ , and  $H_8$ , respectively) for means of successive measurements of the same sample. Finally, tests are provided of the null hypotheses that there is no difference in trend in two series of means: ( $H_9$ ) when each mean in each series is based on an independent sample, ( $H_{10}$ ) when each pair of corresponding means is based on an independent sample, ( $H_{11}$ ) when each series of means is based on an independent sample, and ( $H_{12}$ ) when both series are based on a single sample.

### *Case 1: Test for Trend in Means Based on Independent Samples*

Suppose that several groups of subjects were originally selected at random from the same population and that one group was subjected to a certain treatment for one hour, another group was given the same treatment for two hours, another for three, etc. For each group, measures of a certain trait ( $Y$ ) were obtained immediately following administration of the treatment. Figure 1 represents the possible outcome of an experiment of this type, the open dots representing the means of the criterion variable for the various groups arranged in order of duration ( $X$ ) of treatment. The problem would of course be the same if  $X$  represented amount, or intensity, or num-

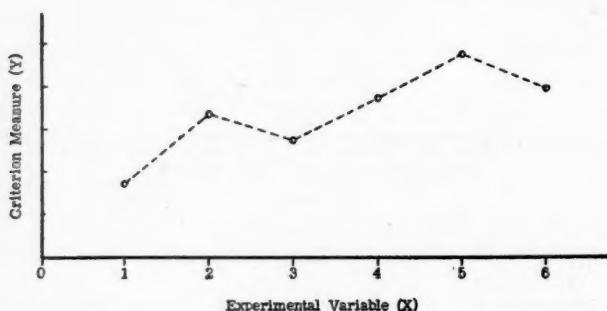


Figure 1

ber of repetitions, etc., of the treatment. (The  $X$  increment or interval need not be uniform.)

The purpose of the experiment is to test the null hypothesis ( $H_1$ ) that the criterion variable ( $Y$ ) is *unaffected* by increases in  $X$  or that the line of population means is a horizontal straight line. The test needed is obviously that provided by the method of simple analyses of variance.

*Procedure:* Assume that the criterion variable is normally distributed and has the same variance for all populations, i.e., for all values of  $X$ . Let  $N$  represent the total number of cases in all samples (groups).

Using the methods of simple analyses of variance, compute the "between groups" or "between intervals" variance ( $\sigma_b^2$ ) and the "within groups" variance ( $\sigma_w^2$ ). Under the hypothesis, each is an unbiased estimate of the common population variance. The hypothesis may then be tested by

$$F = \sigma_b^2 / \sigma_w^2$$

for which the degrees of freedom are  $(k - 1)$  and  $(N - k)$ .

*Note:* Careful consideration should be given in this and the following tests, to the assumption that the population variance (of  $Y$ ) is the same for all values of  $X$ . This is sure to be an unsound assumption if the line of means begins at, or at any point closely approaches, the base line ( $Y = 0$ ) and if negative values of  $Y$  are impossible. In such cases, however, it may sometimes be possible to test the hypothesis along only that part of the  $X$ -scale for which the  $Y$ -variance may be considered fairly constant, disregarding the rest of the data.

*Case 2: Test for Trend in Means Based on a Single Sample*

The experimental design is similar to that in Case 1, except that only one sample of subjects is employed— $k$  successive measures of the criterion, corresponding to the successive values of  $X$ , being obtained for each subject. The purpose in this case is to test the hypothesis ( $H_2$ ) that, for the population involved, the increases in  $X$  are not accompanied by any changes in the  $Y$ -means, or that the regression line of population means is a horizontal straight line. Figure 1 might again represent the experimental results, except that in this case all of the means are based on the same subjects.

*Procedure:* Let  $n$  represent the number of subjects, let  $Y_i$  represent any measure of the criterion variable for the  $i$ -th value of  $X$ , and let  $M_i$  represent the mean of the  $Y_i$ 's.

Consider the individual  $Y$ -measures as arranged in a table in which the  $k$  columns correspond to the various values of  $X$ , and the  $n$  rows correspond to the individual subjects. Using the methods of analysis of variance, compute the "between columns" or "between intervals" variance ( $\sigma_i^2$ ) and the "intervals  $\times$  subjects interaction" variance ( $\sigma_{i \times s}^2$ ) for this table. The hypothesis may then be tested by

$$F = \sigma_i^2 / \sigma_{i \times s}^2$$

for which the degrees of freedom are  $(k - 1)$  and  $(k - 1)(n - 1)$ .

This test involves the important assumptions that, for the entire population sampled, the regression lines for all subjects have zero slope, or are all parallel, and that deviations from individual linear regression are normally distributed and of equal variance for all individuals. Under these assumptions, both  $\sigma_i^2$  and  $\sigma_{i \times s}^2$  are unbiased estimates of this common variance if the hypothesis is true.

*Case 3: Test for Linearity of Means Based on Independent Samples*

In an experiment designed like that in Case 1, it may be desired to test the hypothesis ( $H_3$ ) that increases in  $X$  are accompanied by proportional changes in  $Y$ , or that there is a *linear* relationship between  $X$  and  $Y$ . The hypothetical line of population means in this case is a straight line, but neither its slope nor its  $Y$ -intercept is specified.

*Procedure:* Assume that  $Y$  is normally distributed and has the same variance for each value of  $X$ .

Consider the data as arranged in a correlation table, in which the  $k$  columns correspond to the various samples, or to the  $X$ -inter-

vals, and the rows correspond to the  $Y$ -intervals. Compute the "within columns" variance ( $\sigma_w^2$ ) for the criterion variable. Compute also the "sum of squares" for "between columns" or "between intervals" ( $ss_1$ ). Then, using the methods of analysis of co-variance,\* compute the total "sum of products" ( $\sum xy$ ), and the total "sum of squares" ( $\sum x^2$ ) for the  $X$ -distribution.

The hypothesis may now be tested† by

$$F = \frac{ss_1 - \frac{(\sum xy)^2}{\sum x^2}}{k-2} \bigg/ \sigma_w^2,$$

for which the degrees of freedom are  $(k-2)$  and  $(N-k)$ .

Since the null hypothesis tested in Case I represents a special case of linear relationship, there would be little point here in testing for linearity until the null hypothesis had first been proved untenable.

#### Case 4: Test for Linearity of Means Based on a Single Sample

The experimental design is like that in Case 2,  $k$  successive measures of  $Y$  being obtained for each individual in a single sample. The purpose is to test the hypothesis ( $H_4$ ) that there is a linear relationship between  $X$  and  $Y$ . (See Case 3.)

*Procedure:* Compute the "interaction" variance ( $\sigma_{1 \times S}^2$ ) as in Case 2. Compute also the "sum of squares" for "between intervals" ( $ss_1$ ).

Using the methods of analysis of co-variance,‡ compute the total "sum of products" ( $\sum xy$ ), and the total "sum of squares" ( $\sum x^2$ ) for the  $X$ -distribution.

The hypothesis may now be tested by

$$F = \frac{ss_1 - \frac{(\sum xy)^2}{\sum x^2}}{k-2} \bigg/ \sigma_{1 \times S}^2,$$

for which the degrees of freedom are  $(k-2)$  and  $(n-1)(k-1)$ .

This test involves the assumptions that all individual regression lines are linear and parallel,§ and that deviations from individual re-

\* Lindquist, E. F. Statistical analysis in educational research, Boston: Houghton-Mifflin, 1940, pp. 180-196.

† Ibid., pp. 235-238.

‡ Ibid., pp. 180-186, 196-203, 235-238.

§ This article (somewhat differently organized), was first used in lithoprinted form in the author's classes in March, 1945, and is now being published in response to numerous requests that it be made more generally available in perma-

gression are normally distributed and of equal variance for all subjects.

*Case 5: Goodness of Fit of Curves Not Derived from the Experimental Data (Independent Samples)*

In an experiment designed like that in Case 1, it may be desired to test an *a priori* hypothesis which specifies the exact value\* of the population mean corresponding to each value of  $X$ , or which completely defines the curve of population means ( $L$  in Fig. 2).

This problem may be stated in general terms as follows: *Given a random sample from each of  $k$  populations, to test the hypothesis ( $H_5$ ) that the successive population means\* have the values  $M_{H1}, \dots, M_{H1} \dots$ , and  $M_{Hk}$ , respectively.*

*Procedure:* Let  $Y_i$  represent a single measure in the  $i$ -th sample, let  $n_i$  be the number of cases in the  $i$ -th sample, and let  $N = \sum n_i$ . The total sum of squared deviations of the individual measures from their respective hypothetical means may then be analyzed into two independent components, as follows:

$$\sum \sum (Y_i - M_{Hi})^2 = \sum n_i (M_i - M_{Hi})^2 + \sum \sum (Y_i - M_i)^2.$$

The first component is computed by finding the weighted sum of squared deviations of the sample means from their respective hypothetical means and may be called the "departure from hypothesis" sum of squares. The second is the sum of squares for "within groups" and is computed as a residual in simple analysis of variance (as in Case 1). The hypothesis may now be tested by means of

$$F = \frac{\sum n_i (M_i - M_{Hi})^2}{k} \bigg/ \frac{\sum \sum (Y_i - M_i)^2}{N - k}.$$

Since the hypothesis is not derived from the experimental data in any way, there is no loss of degrees of freedom in the variance esti-

\* It does not matter how one arrives at these values, so long as they are not derived from the experimental data. They may be separately and arbitrarily set, or derived from a regression formula, or read from a graph which has been fitted by inspection to the means derived in a previous experiment, or obtained in any other fashion.

nent form. H. W. Alexander has since published an article on "A General Test for Trend," *Psychological Bulletin*, 1946, 43, 533-557, which contains tests for the hypotheses here designated as  $H_2$  and  $H_4$ . Alexander's tests, while considerably more complex, are superior in that they provide for the possibility of individual differences in regression. Where there is no good reason to suspect heterogeneous regression, the simpler tests here suggested may perhaps be safely employed.

mate based on the "departure from hypothesis" sum of squares. Hence, the d. f. for this  $F$  are  $k$  and  $(N - k)$ . (The denominator is the  $\sigma_w^2$  of Case 1.)

This test of  $H_5$  has the serious weakness that it does not take adequately into consideration either the *direction* or the *pattern* of the discrepancies. In Figure 2, for example, in which  $d_2 = d'_2$  and  $d_3 = d'_3$ , this procedure would have the same outcome if the sample means fell at

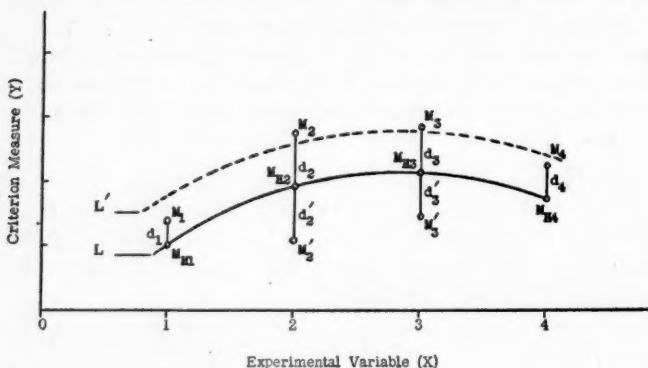


Figure 2

$M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  as if they fell at  $M_1$ ,  $M'_2$ ,  $M'_3$ , and  $M_4$ . However, it obviously is more difficult to account for  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  jointly as chance deviations from line  $L$  than it is to so account for  $M_1$ ,  $M'_2$ ,  $M'_3$ , and  $M_4$ , since, if  $L$  is a true hypothesis, one would expect some of the sample means to fall on either side of  $L$ , instead of all on one side. On the other hand, the *pattern* formed by  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  is much more consistent with the form of the line  $L$  than is the pattern formed by  $M_1$ ,  $M'_2$ ,  $M'_3$ , and  $M_4$ . If  $L$  were shifted upward to the position  $L'$ , for instance, it might represent a very tenable hypothesis as far as  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  are concerned. (In other words, if the hypothesis ( $H_5$ ) is of the form  $\bar{Y} = f(x) + C$ , it is possible that the  $f(x)$  correctly describes the *pattern* of the population means, but that the vertical placement of the pattern is not correctly given by  $C$ .)

It is thus evident that the  $F$ -test suggested above might lead to the rejection of an hypothesis which is very acceptable as far as *pattern* alone is concerned. What is needed, then, is a separate test concerned only with the hypothesis ( $H_{5a}$ ) that the *pattern* of population means as specified by  $H_5$  is true. If the hypothetical pattern is ten-

able, still another test will be needed of the further hypothesis ( $H_{sb}$ ) that the vertical position of that pattern as specified by  $H_s$  is true also.

*Procedure (a):* The desired tests may be based on the two components of the sum of squares for "departure from hypothesis"  $[\sum n_i (M_i - M_{Hi})^2]$ , one of which is due to the "departure from pattern," the other to the "vertical displacement" of that pattern. The sum of squares for "vertical displacement" is computed by

$$[\sum n_i (M_i - M_{Hi})]^2 / N,$$

while that due to "departure from pattern" may be obtained as a residual from the "departure from hypothesis" sum of squares.

The test of  $H_{sa}$  (or the test for "departure from pattern") may then be based on

$$F = \frac{\sum n_i (M_i - M_{Hi})^2 - \frac{1}{N} [\sum n_i (M_i - M_{Hi})]^2}{k - 1} \bigg/ \sigma_w^2,$$

in which  $\sigma_w^2$  is computed as before. The d. f. for this  $F$  are  $(k - 1)$  and  $(N - k)$ .

*Procedure (b):* If the preceding test reveals that the hypothetical pattern is untenable, there is of course no point in going further. However, if  $H_{sa}$  is shown to be tenable,  $H_s$  might still be false with reference to the vertical position of the pattern. The test for  $H_{sb}$  (or the test for "vertical displacement") may then be based on

$$F = \frac{1}{N} [\sum n_i (M_i - M_{Hi})]^2 / \sigma_w^2,$$

for which the d. f. are 1 and  $(N - k)$ . This  $F$ -test is the equivalent of a simple  $t$ -test of the significance of the deviation of the general mean ( $M = \sum n_i M_i / N$ ) from the hypothetical  $M_H = \sum n_i M_{Hi} / N$ .

It should be emphasized that the test for  $H_{sb}$  should never be made unless  $H_{sa}$  has first been proved tenable. If either  $H_{sa}$  or  $H_{sb}$  proves untenable,  $H_s$  is untenable also. The tests for  $H_{sa}$  and  $H_{sb}$  together clearly constitute a better test of  $H_s$  than does the over-all test (for departure from hypothesis) alone.

#### *Case 6: Goodness of Fit of Curves Not Derived from Experimental Data (Single Sample)*

In an experiment designed like that in Case 2, it may be desired to test the hypothesis ( $H_e$ ) that the population means have certain

specified values,  $M_{H1}, \dots, M_{Hi}, \dots$ , and  $M_{Hk}$ , respectively. As in Case 5, it is necessary to apply two separate tests, one of the hypothesis ( $H_{sa}$ ) that the *pattern* of the hypothetical means is correct, the other of the hypothesis ( $H_{sb}$ ) that the vertical placement of that pattern is correct. (Since these tests together will constitute an adequate test of  $H_s$ , there is no need for an over-all test like that for  $H_5$ , although a similar test could readily be supplied.)

*Procedure (a)*: In the  $k \times n$  table referred to in Case 2, compute the "intervals  $\times$  subjects" interaction variance ( $\sigma_{Is}^2$ ). For the same table, compute also

$$\frac{n \sum (M_i - M_{Hi})^2 - \frac{n}{k} [\sum (M_i - M_{Hi})]^2}{k - 1},$$

which corresponds to the "departure from pattern" variance in Case 5.

The hypothesis ( $H_{sa}$ ) may now be tested by

$$F = \frac{n \sum (M_i - M_{Hi})^2 - \frac{n}{k} [\sum (M_i - M_{Hi})]^2}{k - 1} \bigg/ \sigma_{Is}^2,$$

for which the d. f. are  $(k - 1)$  and  $(n - 1)(k - 1)$ .

The assumptions involved are that the individual regression lines are all parallel and that deviations from individual regression are normally distributed and of equal variance for all subjects.

*Procedure (b)*: Again consider the  $Y$  measures as arranged in a table of  $k$  columns ( $X$ -intervals) and  $n$  rows (subjects) as in Case 2. Compute the "between subjects" variance ( $\sigma_s^2$ ) for this table.

If  $H_{sa}$  has been proved tenable,  $H_{sb}$  may then be tested by

$$F = \frac{\frac{n}{k} [\sum (M_i - M_{Hi})]^2}{\sigma_s^2},$$

for which the d. f. are 1 and  $(n - 1)$ .

The test for  $H_{sb}$  may be regarded as a test of the significance of  $(M - M_H)$ , in which  $M = \sum M_i/k$  is the general mean, and  $M_H = \sum M_{Hi}/k$ . The unit of sampling, so far as  $M$  is concerned, is the individual subject, rather than the individual observation. Hence the appropriate error term in this case is  $\sigma_s^2$  rather than  $\sigma_{Is}^2$ .

The assumptions involved are the same as in testing  $H_{sa}$ .

*Case 7: Goodness of Fit of Curves Fitted to the Experimental Means (Independent Samples)*

In an experiment designed like that in Case 1, it may be desired to test an hypothesis ( $H_7$ ) represented by a curved line which has been fitted to the experimental means. The procedure in this case would be exactly like that in testing  $H_{5a}$  (Case 5), except that the d. f. for the numerator of  $F$  would be  $k$  less the number of constants in the regression formula that were derived from the experimental means. For example, if  $Y = a \log X + \bar{b}$  had been fitted\* to the means, the d. f. for the numerator of  $F$  would be  $(k - 2)$ ; if  $\bar{Y} = aX^3 + bX^2 + cX + d$  had been fitted, the d. f. would be  $(k - 4)$ , etc.

In the case of a fitted regression line there would of course be no question of vertical displacement and hence no need for a test like that for  $H_{5b}$ .

*Case 8: Goodness of Fit of Curves Fitted to the Experimental Means (Single Sample)*

In an experiment like that in Case 6, to test an hypothesis ( $H_8$ ) represented by a curved regression line which has been fitted to the experimental means, one would proceed just as in testing  $H_{5a}$ , except that the d. f. for the numerator of  $F$  would be  $k$  less the number of constants in the regression formula derived from the experimental means.

*Case 9: Significance of Trend Differences: Successive Means Based on Independent Samples for Each Trend Separately*

The experiment in this case may be considered as consisting in effect of two parallel experiments of the type in Case 1,  $k$  samples being selected from each of two populations ( $A$  and  $B$ ). The results might be as represented in Figure 3, the solid line joining the observed  $Y$ -means for the samples from population  $A$ , the broken line joining the means of the samples from population  $B$ .

The hypothesis ( $H_9$ ) to be tested is that the two population means coincide for each value of  $X$ . In other words, it is desired to know

\*For methods of fitting curved regression lines, see Fisher, R. A. Statistical methods for research workers, Edinburgh and London: Oliver and Boyd, 1938, Secs. 27-29, pp. 148-177 and Snedecor, George W. Statistical methods, Ames, Iowa: Iowa State College Press, 1940, Chapter 14 (Curvilinear regression), pp. 308-335. The methods there described are for fitting curved regression lines to individual observations, but of course are applicable to means as well. Being concerned with individual observations, however, these discussions do not suggest any tests based upon within-groups variance, such as are necessary if the tests are to be regarded as tests of adequacy of fit.

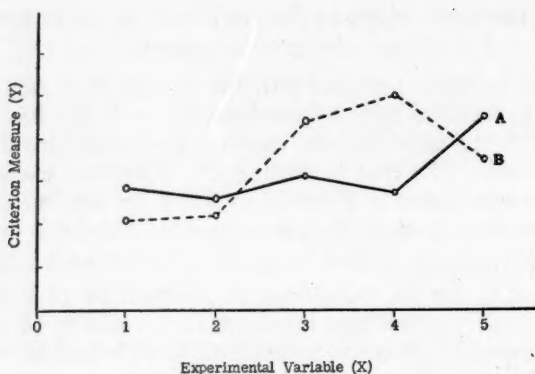


Figure 3

if the differences in the paired sample means can reasonably be jointly attributed entirely to random sampling fluctuations.

The problem would of course be the same if two sets of  $k$  samples were originally selected from the same population, and  $A$  and  $B$  represented different treatments, or different conditions under which the same treatment is administered.

As in Case 5, two separate tests are required.

*Procedure (a):* Consider the data as arranged in a table of 2 (treatments) columns and  $k$  ( $X$ -intervals) rows. Compute the "between treatments" ( $\sigma^2_T$ ), the "treatments  $\times$  intervals interaction" ( $\sigma^2_{T \times I}$ ), and the "within cells" ( $\sigma^2_w$ ) variances for this table. Assume that the  $Y$  measures are normally distributed and have the same variance for all values of  $X$  and for both treatments.

One aspect of  $H_0$  may first be tested by

$$F = \sigma^2_{T \times I} / \sigma^2_w,$$

for which the degrees of freedom are  $(k - 1)$  and  $(N - 2k)$ ,  $N$  being the total number of  $Y$  measures for both sets of samples ( $N = \sum n_{iA} + \sum n_{iB}$ ). This  $F$  tests the hypothesis ( $H_{0a}$ ) that the difference in the means of populations  $A$  and  $B$  is the same for all values of  $X$ . (It does not test the hypothesis that these differences are zero, but only that the successive means of population  $A$  form the same pattern as those of population  $B$ . According to this hypothesis, the two lines of population means would either coincide or lie "parallel" to one another.)

*Procedure (b):* If  $H_{0a}$  proves tenable, another aspect of  $H_0$  may be tested by

$$F = \sigma_T^2 / \sigma_w^2,$$

for which the degrees of freedom are 1 and  $(N - 2k)$ . This tests the hypothesis ( $H_{0b}$ ) that the mean of the successive differences in population means is equal to zero. However, this mean could equal zero even though  $H_{0a}$  were false; hence the test suggested in (a) preceding should be applied first. If either of these  $F$ 's is significant, the hypothesis ( $H_0$ ) is untenable.

*Case 10: Significance of Trend Differences: Successive Mean Differences Based on Independent Samples*

The experiment in this case is designed as in Case 1, except that for each value of  $X$  two criterion measures are secured for each subject, one under Treatment A and one under Treatment B. A different and independent sample, however, is used for each value of  $X$ . The results may be represented as in Figure 3. The hypothesis ( $H_{10}$ ) to be tested is that the population values of the treatment means are identical for each value of  $X$ . The two aspects of this hypothesis are tested separately.

*Procedure (a):* The first hypothesis ( $H_{10a}$ ) to be tested is that the difference in population values of the A and B means is the same (not necessarily zero) for all values of  $X$ .

For each value of  $X$  separately, consider the data as arranged in a table of 2 (treatments) columns and  $n_i$  (subjects) rows. For each such table compute the "treatments  $\times$  subjects" interaction sum of squares ( $ss_{T \times S_i}$ ). Pool these sums of squares for all values of  $X$  and divide by the sum of the corresponding degrees of freedom. The result will be the "error" variance ( $\sigma_{\text{error}}^2$ ) needed in this test.

$$\sigma_{\text{error}}^2 = \frac{\sum ss_{T \times S_i}}{N - k},$$

in which  $N$  is the total number of subjects ( $N = \sum n_i$ ), not the total number of observations, which is  $2N$ . Now consider the data for the entire experiment as arranged in a table of 2 (treatments) columns and  $k$  ( $X$ -intervals) rows. For this table, compute the "treatments  $\times$  intervals" interaction variance ( $\sigma_{T \times I}^2$ ) and the "treatments" variance ( $\sigma_T^2$ ).

The hypothesis ( $H_{10a}$ ) may now be tested by

$$F = \sigma_{T \times I}^2 / \sigma_{\text{error}}^2,$$

with  $(k - 1)$  and  $(N - k)$  d. f.

*Procedure (b):* On the assumption that  $H_{10a}$  is true, one may next test the further hypothesis ( $H_{10b}$ ) that the mean of the successive differences in population means is zero. (This is equivalent to testing the significance of the difference in the general means for "treatments.")

This hypothesis is tested by

$$F = \sigma_T^2 / \sigma_{\text{error}}^2,$$

with 1 and  $(N - k)$  d. f.

If either  $H_{10a}$  or  $H_{10b}$  is untenable,  $H_{10}$  is untenable also.

*Case 11: Significance of Trend Differences: Successive Means  
Based on Single Sample for Each Trend Separately*

An important type of experiment is that which consists, in effect, of two parallel experiments of the type illustrated in Case 2. That is,  $k$  successive measures of the criterion are obtained for each individual in each of two independent samples, each sample being drawn from a different population (or each given a different treatment, or given the same treatment under different conditions). Figure 3 could again represent the experimental results, with the difference that in this case all of the  $A$  means are based on the *same* individuals (instead of on  $k$  independent samples), and that all the  $B$  means are similarly based on a single sample (independent of the  $A$  sample).

The hypothesis to be tested ( $H_{11}$ ) is that the population means coincide for each value of  $X$ .

Two separate tests, corresponding in purpose to those described in Cases 9 and 10, should be applied, as follows:

*Procedure (a):* The first step is to test the more specific hypothesis ( $H_{11a}$ ) that the difference in the  $A$  and  $B$  population means is the *same* (not necessarily zero) for all values of  $X$ .

For this purpose, first regard all of the  $Y$  measures as arranged in a table of 2 (treatments) columns and  $k$  ( $X$ -intervals) rows. Then compute the "treatments  $\times$  intervals interaction" variance ( $\sigma_{T \times I}^2$ ) for this table.

Now consider the  $A$ -data alone as arranged in a table of  $k$  rows and  $n_A$  subjects, and compute the "intervals  $\times$  subjects" sum of squares ( $ss_{I \times S(A)}$ ) for this table. In the same manner, compute ( $ss_{I \times S(B)}$ ) for the  $B$ -data alone. Then pool these sums of squares and divide by

the sum of the corresponding d. f. to secure the "error" variance ( $\sigma_{\text{error}}^2$ ).

$$\sigma_{\text{error}}^2 = \frac{SS_{I \times S(A)} + SS_{I \times S(B)}}{(n_A + n_B - 2)(k - 1)}.$$

This involves the important assumption that the individual regression lines are parallel for all members of both populations and that the deviations from individual regressions are normally distributed and of equal variance for all members of both populations.

The hypothesis ( $H_{11a}$ ) may then be tested by

$$F = \sigma_{T \times I}^2 / \sigma_{\text{error}}^2,$$

for which the degrees of freedom are  $(k - 1)$  and  $(n_A + n_B - 2)(k - 1)$ .

*Procedure (b):* If  $H_{11a}$  is proved tenable, the next step is to test the hypothesis ( $H_{11b}$ ) that the mean of the differences in population means for the various values of  $X$  is zero. It should be apparent that this is equivalent to testing the significance of the difference ( $M_A - M_B$ ) in the general means of the two samples.

For this purpose, consider the data for the  $A$ -sample above as arranged in a  $k \times n_A$  table, and compute the "sum of squares for between subjects" ( $SS_{S(A)}$ ) for this table. Treat the  $B$ -data similarly to compute  $SS_{S(B)}$ .

If  $n_A$  and  $n_B$  are small, the significance of ( $M_A - M_B$ ) may be tested by

$$t = (M_A - M_B) / \sqrt{\frac{SS_{S(A)} + SS_{S(B)}}{k(n_A + n_B - 2)} \cdot \left(\frac{1}{n_A} + \frac{1}{n_B}\right)},$$

for which the number of degrees of freedom is  $(n_A + n_B - 2)$ .

If  $n_A$  and  $n_B$  are large, the significance of ( $M_A - M_B$ ) may be tested by the normal deviate (critical ratio)

$$C.R. = (M_A - M_B) / \sqrt{\frac{SS_{S(A)}}{k(n_A - 1)} + \frac{SS_{S(B)}}{k(n_B - 1)}},$$

avoiding the necessity of any assumption of equal "between subjects" variances for the two populations.

If either  $H_{11a}$  or  $H_{11b}$  must be rejected, then, of course,  $H_{11}$  must be rejected also.

*Case 12: Significance of Trend Differences: Successive Means  
Differences Based on a Single Sample*

The design of the experiment in this case is like that of Case 11, except that both treatments are administered to the same sample of subjects, instead of one to each of two independent samples. Again the purpose is to test the hypothesis ( $H_{12}$ ) that the population means for the two treatments are identical for all values of  $X$  or that there is no difference in the two treatments. Figure 3 could again represent the experimental results—all means, for both  $A$  and  $B$ , in this case being based on the same individuals.

*Procedure (a):* In terms of the methods of analysis of variance, this is a factorial design involving the three factors: "treatments," "intervals," and "subjects." As in the preceding cases, two separate tests are needed.

The first step is to test the hypothesis ( $H_{12a}$ ) that the differences in  $A$  and  $B$  means for the various values of  $X$  all have the same population values.

Using the methods of analysis of variance for factorial designs, compute the "treatments" variance ( $\sigma^2_T$ ), the interaction variance for "treatments  $\times$  intervals" ( $\sigma^2_{T \times I}$ ), the interaction variance for "treatments  $\times$  subjects" ( $\sigma^2_{T \times S}$ ), and the higher order interaction for "treatments  $\times$  intervals  $\times$  subjects" ( $\sigma^2_{T \times I \times S}$ ).

The hypothesis may now be tested by

$$F = \sigma^2_{T \times I} / \sigma^2_{T \times I \times S},$$

for which the degrees of freedom are  $(k - 1)$  and  $(k - 1)(n - 1)$ .

*Procedure (b):* The next step is to test the hypothesis ( $H_{12b}$ ) that the mean of the differences in  $A$  and  $B$  means is zero for the entire population.

This hypothesis would not be tested unless  $H_{12a}$  had first been shown to be tenable. On the assumption that there is no real  $T \times I$  interaction, the hypothesis  $H_{12b}$  may be tested by

$$F = \sigma^2_T / \sigma^2_{T \times S},$$

for which the degrees of freedom are 1 and  $(n - 1)$ .

## THE ESTIMATION OF RELIABILITY WHEN SEVERAL TRIALS ARE AVAILABLE\*

HOWARD W. ALEXANDER

ADRIAN COLLEGE

In certain types of experiment it frequently happens that we have a block of data consisting of several trials on the same individuals. Using the methods of estimation provided by the analysis of variance, estimates of reliability are derived for this case, and the conditions under which each is valid are discussed. Various relations between these estimates and the product-moment coefficient of correlation are obtained.

It frequently happens in educational, psychological or biological research that it is possible to obtain repeated measurements on a limited number of subjects over a considerable period of time. Suppose, for example, that we have 12 experimental subjects maintained under constant "standard" conditions, and that we obtain 6 trials of a certain measure (say an intellective test) spaced two days apart. From the block of data thus provided, consisting of the scores of 12 individuals on six trials, can we estimate the reliability of the test? How will such an estimate compare with other measures of reliability obtained from just two trials?

In many types of biological experiment it is difficult to obtain large numbers of subjects. Thus estimates of reliability based on the accepted methods may suffer the serious disadvantage of being based on small "degrees of freedom." It is of considerable importance to know whether, and to what extent, this handicap can be compensated for by a larger number of trials. The present paper will discuss methods of estimating reliability when a relatively large number of trials is available on a relatively small group of subjects.

In Part I (sections 1 to 8) methods of reliability estimation based

\* From the Laboratory of Physiological Hygiene, University of Minnesota. This work was supported in part under the terms of a contract between the Regents of the University of Minnesota and the Office of Scientific Research and Development. Important financial assistance was also provided by the Nutrition Foundation, Inc.; The U. S. Sugar Cane Refiners Association, N. Y.; the Corn Industries Research Foundation, N. Y.; Swift and Co., Chicago; the National Confectioners' Association; the National Dairy Council, Chicago; and the Graduate Medical Research Fund, University of Minnesota. The author is indebted for assistance and encouragement to all the members of the staff of the Laboratory of Physiological Hygiene, and especially to Mr. Harold Guetzkow and Mr. Richard Seymour.

on the analysis of variance are developed. In this development, the case of two trials is used in order that the estimates derived may be compared with the more usual measures of reliability. Emphasis is placed on the use of estimates of population variance as the fundamental entities from which estimates of reliability may be constructed.

Part II (sections 9 to 13) deals with the case of more than two trials. A general test for trend is outlined. Equations for the estimation of variance when a component of linear trend is present are developed. Finally, the estimates of reliability are studied from the standpoint of average correlation and in the light of maximum likelihood.

### *Part I*

#### *1. The concept of reliability*

The concept of reliability has been very adequately discussed in a publication by R. W. B. Jackson (5), in which a rather complete bibliography is provided. The reliability of a test or measure is concerned with its *accuracy*. Thus, some estimate of the error of measurement of the test will be included as one component of any measure of reliability. "Error of measurement" may be taken to mean, for example, the failure of the individual to repeat his performance under conditions as nearly identical as possible. For a single individual, this may be measured by the variance of his scores; for a group we may use the average variance for all individuals. Under other circumstances, the error may be measured by the deviation of the individual from the average performance of the group, in which case it is essentially a measure of the inconsistency of the performance of the group. In this case, the appropriate quantitative measure of error is readily obtained by the methods of analysis of variance from the quantity known as the "interaction mean square."

The measure of error may be reported in absolute units. Or it may be expressed as a percentage of the mean of the sample from which it was derived, a procedure common in the physical and biological sciences. Or, again, it may be compared with the inter-individual variability; this alternative gives rise to the product-moment coefficient of correlation, to the intraclass correlation, and to certain other measures.

#### *2. The basic data for reliability estimation*

Consider an aggregation of measurements which have been repeated  $k$  times on  $n$  individuals. These measurements may be arrayed as in Table 1.

TABLE 1

	Trials				$k$
	1	2	3		
1	$X_{11}$	$X_{12}$	$X_{13}$	...	$X_{1k}$
2	$X_{21}$	$X_{22}$	$X_{23}$	...	$X_{2k}$
3	$X_{31}$	$X_{32}$	$X_{33}$	...	$X_{3k}$
Indi-					
duals					
$n$	$X_{n1}$	$X_{n2}$	$X_{n3}$		$X_{nk}$

If the subscript  $j = 1, 2, 3, \dots, n$  designates the individual, and the subscript  $h = 1, 2, 3, \dots, k$  designates the trial or repetition, then the score for any individual on any trial may be designated by  $X_{jh}$ .

In some cases the repetitions or trials are obtained over time; in other cases (as in certain kinds of psychological tests) duplicate measurements are obtained by splitting the odd and even items of a test into two groups. The term "individual" can be applied to objects as well as persons. For example, in a geological investigation we might consider the "individuals" to be different kinds of gravel; in this instance we might use a large number of samples of the same kind of gravel, rather than repeated measurements on the same sample. In many kinds of biological experiment, however, it is more convenient to repeat the measurements frequently on the same individual.

As will be brought out in later sections, the choice of a suitable measure of reliability is dependent upon the presence or absence of *trend*. We may define trend, in a table of individuals by trials such as Table 1, to be a fluctuation of the sample mean from trial to trial that is too large to be attributed to the error of measurement. Jackson (4) has used the term "trial effect" or "practice effect" for what we have named trend, and uses methods of analysis of variance to test for it.

### 3. The test for trend in the case of two trials

Throughout the remainder of Part I we shall be concerned primarily with the case of two trials, which may be displayed as in Table 2.

TABLE 2

	Trials		Difference
	1	2	
1	$X_{11}$	$X_{12}$	$d_1$
2	$X_{21}$	$X_{22}$	$d_2$
Indi-			
duals			
$n$	$X_{n1}$	$X_{n2}$	$d_n$

We have supplied a difference column,  $d_j = X_{j1} - X_{j2}$ . In the case of two trials, a simple test for trend is provided by the  $t$ -test for paired variates, which is a test of the hypothesis that the mean of the first trial is the same as the mean of the second trial. Numerically, it is given by the formula

$$t = \frac{\sqrt{n-1} (\Sigma d)}{\sqrt{n \Sigma d^2 - (\Sigma d)^2}}. \quad (1)$$

The value of  $t$  is to be compared with tabulated values for  $n - 1$  df (degrees of freedom). However, it is preferable to use the equivalent  $F$ -test, since in preparing to calculate  $F$  we find other quantities which are needed in the estimation of reliability.

The calculations required to set up the table of analysis of variance are described in numerous texts (8, Chap. 11). The features in which we are most interested are illustrated in the two fictitious examples in Table 3.

TABLE 3

Example 1				Example 2			
$X$	$Y$	$X + Y$	$X - Y$	$X$	$Y$	$X + Y$	$X - Y$
64	67	131	-3	64	57	121	7
27	28	55	-1	27	18	45	9
93	92	185	1	93	82	175	11
45	50	95	-5	45	40	85	5
84	90	174	-6	84	80	164	4
79	76	155	3	79	66	145	13
37	56	93	-19	37	46	83	-9
38	33	71	5	38	23	61	15
96	73	169	23	96	63	159	33
62	60	122	2	62	50	112	12
19	10	29	9	19	0	19	19
16	23	39	-7	16	13	29	3
660	658	1318	2	660	538	1198	122

Note that the tables have been constructed in such a way that the column of  $X$ -values is the same in both examples, while the  $Y$ -values in Example 2 are 10 less than the corresponding  $Y$ -values in Example 1. The two analyses are as given in Table 4.

TABLE 4

Example 1			Example 2		
Source of variation	df	Mean square	df	Mean square	
Between individuals	11	$V_i = 1,476.076$	11	$V_i = 1,476.076$	
Between trials	1	$V_t = .167$	1	$V_t = 620.167$	
Interaction	11	$V_{it} = 51.348$	11	$V_{it} = 51.348$	

Only the between-trials mean square has been affected by the adjustment of the  $Y$ -values in Example 2. The  $F$ -test for the presence of trend is, for the case of two trials, equivalent to the  $t$ -test already given, and is provided by the ratio  $F_t = V_t/V_{it}$ . For the illustrative examples we find:

Example 1:  $F_t = .003$     Example 2:  $F_t = 12.08$

Reference to Snedecor's table of  $F$  (7) shows the first of these two values is non-significant while the second is significant (that is, the value of  $F_t = 12.08$  lies beyond the 1 per cent value of  $F$ , for 1 and 11 df). We conclude that there is a significant trend in the second case, but not in the first. The trend has been introduced by the uniform displacement of the  $Y$ -values in the second example.

#### 4. *Components of variation and estimates of variance*

We now introduce a set of assumptions basic to the development of the various measures of reliability. Returning to the case of  $k$  trials, we postulate that each of the observed values in a table such as Table 1 can be expressed as the sum of four parts:

$$X_{jk} = M + I_j + T_k + R_{jk},$$

where

$M$  = the population mean, constant for all individuals on all trials,

$I_j$  = a component characteristic of the particular individual, but constant over trials.

$T_k$  = a component characteristic of the particular trial, but constant over individuals,

$R_{jk}$  = the residual, which does not vary systematically either with individual or with trial.

We assume that the components  $I_j$ ,  $T_k$ , and  $R_{jk}$  are normally distributed and uncorrelated in the parent population, with mean zero and variances  $\sigma_i^2$ ,  $\sigma_t^2$ , and  $\sigma_e^2$ , respectively. These three population variances will be referred to as "between individuals variance," "between trials variance," and "error variance," respectively. For the validity of the methods which will be developed, it is necessary to assume, further, that the individuals constitute a random sample from the population of such individuals and that the trials constitute a random sample from the population of all similar trials. A more complete discussion of several of these assumptions is given in (2), which also contains a useful bibliography.

It is natural to ask whether a consecutive series of trials can ever be independent, or whether a given set of trials can ever be regarded

as a random sample from the population of trials. An experimental example will be described in some detail to illustrate these possibilities. Twelve normal young men were maintained on a diet as constant in its principal components as was consistent with reasonable variety. On six occasions several days apart electrocardiograms were obtained before and after the noon meal to determine the effect of the meal on the various electrocardiographic measures. Since all the external factors were kept as constant as possible during the experimental period, any trend that might appear would properly be described as random, in the sense that it was uncorrelated with any controllable external factor. A statistical test for trend was applied to one of the electrocardiographic functions known as "sigma  $T$ ", and it was found that while there was no linear or cumulative trend, there was a significant trend common to the group, presumably attributable to the differences between the six meals. The detailed analysis is presented in (1). It is quite clear that the meals were chosen randomly, since the meal sequence was planned to be as homogeneous as possible, and there was no attention paid to the possible effect of any dietary components on electrocardiographic performance. Such random influence of significant factors can doubtless occur in many situations.

While the population variances cannot be calculated from the information in the sample, unbiased estimates  $s_i^2$ ,  $s_t^2$ , and  $s_e^2$  of  $\sigma_i^2$ ,  $\sigma_t^2$ , and  $\sigma_e^2$ , respectively, can be obtained from the equations

$$\begin{aligned} V_i &= s_e^2 + k s_i^2 \\ V_t &= s_e^2 + n s_i^2 \\ V_{it} &= s_e^2 \end{aligned} \quad (2)$$

(see, for example, 7, section 10.6).

The  $F$ -test,  $F_t = V_t/V_{it}$ , may be used to test for trend in this case as in the simple case of two trials. Note that

$$F_t = V_t/V_{it} = \frac{s_e^2 + n s_i^2}{s_e^2} = 1 + n \left( \frac{s_i^2}{s_e^2} \right),$$

which shows that when  $F_t = 1$ ,  $s_i^2 = 0$ . This makes plausible the fact that  $F_t$  can also be regarded as a test of whether  $s_i^2$  is significantly different from zero. If  $s_i^2$  is not significantly different from zero, the component  $T_h$  is unnecessary, since there is no trial component which can be distinguished from error variation. Under these circumstances we replace  $V_t$  and  $V_{it}$ , which are both estimates of error, by the pooled estimate known as the "within individuals" mean square,  $V_{wi}$ :

$$V_{wi} = \frac{1}{n} (V_t + (n-1)V_{it}),$$

with  $n(k-1)$  df. For Example 1 we obtain the analysis:

Source of variation	df	Mean square
Between individuals	11	$V_i = 1,476.076$
Within individuals	12	$V_{wi} = 47.083$

Since the hypothesis that  $\sigma_i^2 = 0$  has been confirmed by the test,  $F_i$ , we drop it from the equations of estimation, and equations (2) are replaced by the following pair of estimation equations:

$$\begin{aligned} V_i &= s'_e{}^2 + ks'_i{}^2 \\ V_{wi} &= s'_e{}^2, \end{aligned} \quad (3)$$

where  $s'_i{}^2$  and  $s'_e{}^2$  are new estimates of  $\sigma_i^2$  and  $\sigma_e^2$ . In the case of Example 1, equations (3) yield  $s'_i{}^2 = 714.496$ ,  $s'_e{}^2 = 47.083$ .

#### 5. Estimates of reliability

All the estimates of reliability that will be developed are ultimately derived from the product-moment coefficient of correlation. Jackson has shown (5, Appendix A, pp. 107-112) that under circumstances which arise frequently estimates of reliability other than the product-moment coefficient are preferable (see sections 7 and 12 below). In the present section we shall develop the purely algebraic properties of four estimates of reliability, the uses of which will be discussed later. Of these, all except the first are given for the general case of  $k$  trials.

#### The product-moment coefficient of correlation

(a) Consider Table 2, with columns of  $X_{j1}$ ,  $X_{j2}$ , and  $d_j$ . Let  $s_1^2$ ,  $s_2^2$ , and  $s_d^2$  be the unbiased estimates of the variances of these three columns:

$$\begin{aligned} s_1^2 &= \frac{n \sum X_{j1}^2 - (\sum X_{j1})^2}{n(n-1)}, & s_2^2 &= \frac{n \sum X_{j2}^2 - (\sum X_{j2})^2}{n(n-1)}, \\ s_d^2 &= \frac{n \sum d_j^2 - (\sum d_j)^2}{n(n-1)}. \end{aligned}$$

It may be shown, by simple algebra, that the following equations hold for the case of two trials:

$$\begin{aligned} V_i &= s_1^2 + s_2^2 - \frac{1}{2} s_d^2, & V_{ii} &= s_e^2 = \frac{1}{2} s_d^2, \\ s_i^2 &= \frac{1}{2} (s_1^2 + s_2^2 - s_d^2). \end{aligned} \quad (4)$$

Finally, the well-known formulas for the product-moment correlation lead to

$$r = \frac{s_1^2 + s_2^2 - s_d^2}{2s_1s_2} = \frac{s_i^2}{s_1s_2}. \quad (5)$$

Using these formulas we find, for both Example 1 and Example 2:

$$s_1^2 = 816.909, \quad s_2^2 = 710.515, \quad s_d^2 = 102.697, \\ s_i^2 = 712.364, \quad r = .9350.$$

Note that the product-moment coefficient is insensitive to the shift in the values on the second trial. It does not reflect the presence or absence of trend.

*Intraclass correlation: symmetrical table estimate.*

(b) R. A. Fisher, in his *Statistical methods for research workers*, (3, 207) makes use of two different forms of the intraclass correlation. The first is estimated from a symmetrical table, using the formula of Harris:

$$k \sum (\bar{x}_p - \bar{x})^2 = (n-1) s^2 [1 + (k-1)r''],$$

where  $r''$  is the symmetrical table estimate of the intraclass correlation,  $\bar{x}$  is the over-all mean,  $\bar{x}_p$  is the mean of the scores for the  $p^{\text{th}}$  individual, and  $s^2$  is an estimate of the total variance. The value of  $r''$  may be expressed in terms of  $V_i$  and  $V_{wi}$ :

$$r'' = \frac{(n-1)V_i - nV_{wi}}{(n-1)V_i + n(k-1)V_{wi}}, \quad (6)$$

or in terms of the ratio  $F'_i = V_i/V_{wi}$ :

$$r'' = \frac{F'_i - n/(n-1)}{F'_i + n(k-1)/(n-1)}. \quad (7)$$

It will be shown later that this estimate is really only appropriate for a case such as Example 1, where no trend is present. For Example 1, we find  $r'' = .9327$ , which is a fairly close estimate of the product-moment correlation,  $r = .9350$ . If we calculate  $r''$  for Example 2, we find  $r'' = .8640$ , which is a serious underestimate of  $r$ . The effect of the presence of trend is to depress the value of  $r''$ .

*Intraclass correlation: unbiased estimate.*

(c) The unbiased estimate of the intraclass correlation, which will be denoted by  $r'$ , takes a very simple form when it is expressed in terms of the estimates  $s'^2_i$  and  $s'^2_e$  of equations (3):

$$r' = \frac{s'^2_i}{s'^2_i + s'^2_e}. \quad (8)$$

This form is susceptible of an intuitive interpretation. The numerator is an estimate of the between-individuals variance, while the denominator is an estimate of the total variance. Thus  $r'$  measures the fraction of the total variance that is attributable to inter-individual variation. More convenient for computational purposes are the expressions

$$r' = \frac{V_i - V_{wi}}{V_i + (k-1)V_{wi}} \quad (9)$$

$$= \frac{F'_i - 1}{F'_i + k - 1}, \quad \text{where } F'_i = V_i/V_{wi}.$$

Again, this estimate is suitable only when there is no trend. However, we present its values for both Example 1 and Example 2:

Example 1:  $r' = .9382$ , Example 2:  $r' = .8746$ .

Note that, as with the previous estimate,  $r'$  is rather close to  $r$  when there is no trend but underestimates  $r$  when trend is present.

*Intraclass correlation adjusted for trend.*

(d) A fourth form of estimate will be useful. It occurs, for the case of two trials, in Jackson's monograph on reliability (6, 110, formula 83). We shall denote by  $r^*$  and describe it as the intraclass correlation adjusted for trend. We may define it initially in terms of the variance estimates of equations (2):

$$r^* = \frac{s_i^2}{s_i^2 + s_e^2} \quad (10)$$

Although this has the same form as equation (8), it must be clearly distinguished from the equation for  $r'$ . The primed estimates of equation (8) are derived from equations (3), for the case without trend. The estimates of equation (10) are the same as those in equations (2), and are for the case with trend. In this case the denominator is not an estimate of the total variance, since the trend component  $s_t^2$  does not appear; it is the "total variance apart from trend."

Some justification for the name given to this estimate may be offered. Suppose that an experiment were designed for the purpose of estimating reliability, using a number of trials. It was intended that performance should be kept steady, so that no trend would be present and the intraclass correlation could be used. Suppose it were found that a certain amount of trial fluctuation had entered. Then  $r^*$  may be regarded as an estimate of the value that would have been obtained if the fluctuation had been avoided.

Equations (10) and (2) lead to simpler computational forms:

$$\begin{aligned} r^* &= \frac{V_i - V_{it}}{V_i + (k-1)V_{it}} \\ &= \frac{F_i - 1}{F_i + k - 1}, \quad \text{where } F_i = V_i/V_{it}. \end{aligned} \quad (11)$$

Note that the above expressions are closely analogous to those in (9).

It may be noted here that  $F'_i = V_i/V_{wi}$  is a test of whether  $r'$  is significantly different from zero, while  $F_i = V_i/V_{it}$  is a test of whether  $r^*$  is significantly different from zero.

For both Example 1 and Example 2 we find

$$F_i = 28.75, \quad r^* = .9328.$$

Thus  $r^*$  in this case provides a fair estimate of  $r = .9350$  and is independent of the presence of trend.

#### 6. Algebraic relations between the estimates

##### (a) Relation between $r$ and $r^*$ .

For the case of two trials,  $r \geq r^*$ . For proof, note that from equations (4) it follows that (for  $k=2$ )

$$V_i - V_{it} = 2s_i^2, \quad V_i + V_{it} = s_1^2 + s_2^2.$$

Thus, for  $k=2$ , we find

$$\begin{aligned} r^* &= \frac{V_i - V_{it}}{V_i + V_{it}} = \frac{2s_i^2}{s_1^2 + s_2^2} \\ &= \frac{s_i^2}{\frac{1}{2}(s_1^2 + s_2^2)}. \end{aligned} \quad (12)$$

The expression in the denominator is the arithmetic mean of the variances of the two trials. On the other hand, in the expression

$$r = s_i^2/s_1s_2$$

of equation (5), the denominator is the geometric mean of the variance of the two trials, while the numerator is the same as in (12). But the identity

$$\frac{1}{2}(s_1^2 + s_2^2) = s_1s_2 + \frac{1}{2}(s_1 - s_2)^2$$

shows that the arithmetic mean of the variance is somewhat greater than the geometric mean, except when the variances are equal, in which case the arithmetic and geometric means are also equal. From this it follows that  $r \geq r^*$ .

(b) *Relation between  $r^*$  and  $r'$ .*

We shall show that, for all values of  $k$ ,

$$\begin{aligned} r^* &> r' && \text{if } F_i > 1, \\ r^* &= r' && \text{if } F_i = 1, \\ r^* &< r' && \text{if } F_i < 1. \end{aligned}$$

We first note the identity

$$F_i = F'_i \left(1 + \frac{F_i - 1}{n}\right) = F'_i + \frac{F'_i}{n} (F_i - 1),$$

which is a consequence of the relations  $F_i = V_i/V_{it}$ ,  $F'_i = V_i/V_{wi}$  and

$$V_{wi} = \frac{1}{n} (V_i + (n-1)V_{it}).$$

Using the above identity, we may write

$$r^* = \frac{F_i - 1}{F_i + k - 1} = \frac{F'_i - 1 + \frac{F'_i}{n} (F_i - 1)}{F'_i + k - 1 + \frac{F'_i}{n} (F_i - 1)}.$$

Comparing this with the expression (9) for  $r'$ , we see that  $r^*$  is obtained from  $r'$  by adding (if  $F_i > 1$ ) the same positive quantity to numerator and denominator of the fraction, which has the effect of increasing its value. Thus, if  $F_i > 1$ ,  $r^* > r'$ . The other relations are proved similarly.

(c) *Relation between  $r'$  and  $r''$ .*

For all values of  $k$ ,  $r' > r''$ . To see this, we note that the expression (7) for  $r''$  may be written

$$r'' = \frac{F_i - 1 - \frac{1}{n-1}}{F_i + k - 1 + \frac{k-1}{n-1}}.$$

Comparing this with (9), we see that  $r''$  is obtained from  $r'$  by subtracting a quantity from the numerator and adding a quantity to the denominator, which has the effect of reducing the value of the fraction. Thus  $r' > r''$ .

7. *Maximum likelihood estimates*

R. W. B. Jackson has shown (6, 107-112) that, for the case of two trials, the maximum likelihood estimate (MLE) of the correlation between two trials depends upon the presence or absence of trend

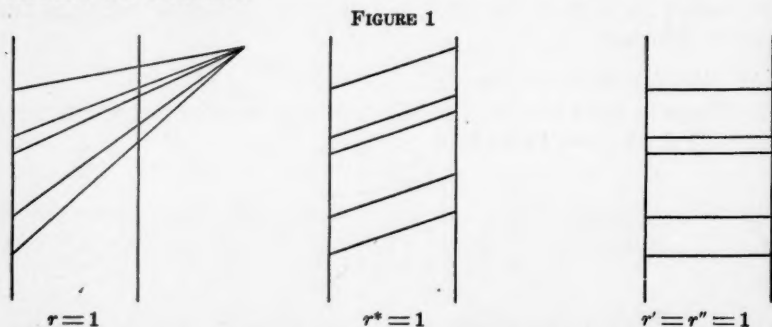
and the equality or inequality of the population values of the variances of the two trials. He distinguishes three cases:

- (a) Trend is present and the variances of the two trials are unequal. In this case, Jackson shows that the MLE is given by the product-moment coefficient of correlation.
- (b) Trend is present and the variances are equal. Here the MLE is shown to be given by an expression algebraically equivalent to  $r^*$  as defined in (11).
- (c) There is no trend and the variances are equal. The MLE is found to be given by the expression (6) for  $r''$ .

The last two of these results will be generalized to the case of  $k$  trials in section 12.

#### 8. A geometrical note

For the case of two trials, an interesting geometrical comparison may be made of the various estimates. Let the two trials be represented by two parallel vertical lines, and let the scores be graphed upon these lines. Then if we join corresponding points on the two lines, we get a series of  $n$  lines, one for each individual. Parenthetically, it may be noted that if  $V_{it} = 0$ , the  $n$  lines will be parallel. What conditions are imposed upon the lines if the different estimates of reliability are put equal to unity? The three parts of Figure 1 illustrate the situation.



The condition  $r = 1$  requires that the  $n$  lines be convergent (which includes parallelism as a special case),  $r^* = 1$  requires that they be parallel, while  $r' = r'' = 1$  requires that they be horizontal. Thus the most stringent condition is  $r' = r'' = 1$ , and the least stringent is  $r = 1$ . Morton Jellinek (7) points out the fact that  $r'$  takes its maximum value only when the values on both trials are identical, offering this as a reason for the adoption of  $r'$  as a measure of reliability. How-

ever, Jackson's result makes it clear that the use of  $r'$  is justified only when no trend is present. The developments below will show that this result holds also for the case of  $k$  trials.

## Part II

### 9. A general test for trend

We must now define more specifically what is to be understood by the term "trend" when we are dealing with more than two trials. The test for trend will be dependent upon the assumptions which are made regarding the distribution of observed values.

Consider, first, the test for trend which results when the basic assumptions are those of section 4. The ratio  $F_t = V_t/V_{it}$  provides, in this case, a test of whether the trial means are different. This is a suitable test for trend under the assumptions above, but some attention must be paid to the restrictive effect of these assumptions.

Consider the implications of the assumption that  $T_h$  is a normally distributed random variable with mean zero. This implies that on any particular trial the mean is just as likely to fall below the population mean as above it. Thus, under this assumption there can be no cumulative trend. These considerations make it clear that the set of assumptions of section 4 are not fulfilled in a situation involving learning, for if later scores tend to be higher, as with learning, the means cannot be randomly distributed.

Another consequence requires emphasis. Suppose that the experiment were repeated a great many times with the same individuals, who will be supposed not to change between repetitions. For each individual on each trial there would thus arise a population of values distributed about a certain mean value  $\tilde{X}_{jh}$ . It is a consequence of our assumptions that if we graph the  $k$  trial values  $\tilde{X}_{jh}$  for each individual against time and join these points by a broken line, then the  $n$  graphs for the individuals of the group will be *parallel*, that is, any one of them may be made to coincide with any other one by vertical displacement.

If we have *a priori* reason for believing that the individual graphs will not be parallel, then the assumptions of section 4 are unsuitable. This constitutes an additional reason for regarding these assumptions as inappropriate for a situation in which learning is present, since the likelihood is that individuals will learn at different rates. However, the experimental example described in section 4 indicates that examples may be encountered in which we have no reason for distrusting the assumptions of section 4.

It is important to be able to test for cumulative as well as for

random trend. There are many types of learning function that could presumably be employed. In most circumstances a linear learning function is probably adequate, and the discussion in the present paper will be confined to the case of linear trend. We shall modify the assumptions of section 4 by postulating that a linear term is present in addition to the components already described. Thus the observed values may be partitioned into five components:

$$X_{jh} = M + I_j + T_h + t_h B_j + R_{jh},$$

where  $M$ ,  $I_j$ ,  $T_h$  and  $R_{jh}$  are as previously defined, and

$B_j$  = a slope term characteristic of the individual,

$t_h$  = the time of the  $h^{\text{th}}$  trial, measured in any convenient units.

For computational convenience, the origin of the time variable is chosen to coincide with the mean value of  $t_h$ , so that  $\sum t_h = 0$ .

The analysis of this case is more fully discussed in (1), where tests for trend are developed for the general case and for the case in which the term  $T_h$  is deleted. In this account only the general case will be considered.

The computational details may best be indicated by starting with the simple analysis of variance such as was used in section 3. We use the notation  $SS_i$ ,  $SS_t$ , and  $SS_{it}$  to denote the sums of squares corresponding to  $V_i$ ,  $V_t$ , and  $V_{it}$ . Thus we have the table:

Source of variation	Sum of squares	df	Mean square
Between individuals	$SS_i$	$n - 1$	$V_i$
Between trials	$SS_t$	$k - 1$	$V_t$
Interaction	$SS_{it}$	$(n - 1)(k - 1)$	$V_{it}$

In addition to these quantities, we require to compute for each individual the sum of products  $P_j = \sum t_h X_{jh}$  (summed over trials). We also compute  $K = \sum t_h^2$  (summed over trials) and the quantities

$$Q = \frac{1}{nK} (\sum P_j)^2 \quad \text{and} \quad R = \frac{1}{K} \sum P_j^2,$$

both summed over individuals.

The analysis yields five mean squares, and the table of analysis of variance is as follows:

Source of variation	Sum of squares	df	Mean square
Error	$SS_{it} + Q - R$	$(n - 1)(k - 2)$	$V_e$
Group deviations from linearity	$SS_i - Q$	$k - 2$	$V_{gd/l}$
Group slope	$Q$	1	$V_{gs}$
Between individual slopes	$R - Q$	$n - 1$	$V_{is}$
Between individual means	$SS_i$	$n - 1$	$V_i$

From this analysis there result two tests for trend:

(a)  $F_{gdf1} = V_{gdf1}/V_e$  is a test of whether the whole group shares a common pattern of variation in addition to the linear component which may vary from individual to individual.

(b)  $F_{gs} = V_{gs}/V_e$  is a test of whether there exists a significant group slope.

As an example, consider the analysis of the work pulse rate for 35 normal young men. Pulses were determined after five minutes on a motor-driven treadmill, on three occasions, with intervals of one week between the trials. The trial means were:

Trial 1	Trial 2	Trial 3
135.6	136.9	132.8

The simple analysis of variance is

Source of variation	Sum of squares	df	Mean square
Between individuals	$SS_i = 16,928.63$	34	$V_i = 497.90$
Between trials	$SS_t = 305.20$	2	$V_t = 152.60$
Interaction	$SS_{it} = 3,240.80$	68	$V_{it} = 47.66$

The auxiliary quantities required for the complete analysis are

$$Q = 145.73, \quad R = 2424.50.$$

Using these, we obtain the analysis

Source of variation	Sum of squares	df	Mean square
Error	962.03	34	$V_e = 28.295$
Group deviations from linearity	159.47	1	$V_{gdf1} = 159.47$
Group slope	145.73	1	$V_{gs} = 145.73$
Between individual slopes	2,278.77	34	$V_{is} = 67.02$
Between individual means	16,928.63	34	$V_i = 497.90$

$$F_{gdf1} = V_{gdf1}/V_e = 5.64, \quad F_{gs} = V_{gs}/V_e = 5.15.$$

Both of these  $F$ -values are significant, so that there is both a significant group slope and a significant variation about the group slope which is shared by the whole group. The existence of significant group slope means that the simple analysis of variance is not valid, since the trial means are not randomly distributed about the grand mean.

#### 10. Estimation when trend is present

When the tests of significance of section 9 show that linear trend is present, then neither of the estimation equations (2) or (3) is

valid. Let us suppose that tests of significance have shown that each term of the equation

$$X_{jh} = M + I_j + T_h + t_h B_j + R_{ij} \quad (13)$$

represents a significant type of variation. Let  $s_i^2$ ,  $s_t^2$ ,  $s_b^2$ , and  $s_r^2$  designate unbiased estimates of the variances of the terms  $I_j$ ,  $T_h$ ,  $B_j$ , and  $R_{jh}$ . The mean values of  $I_j$ ,  $T_h$ , and  $R_{jh}$  are assumed to be zero. Then we obtain the estimation equations

$$\begin{aligned} V_e &= s_r^2 \\ V_{gdf1} &= s_r^2 + n s_i^2 \\ V_{gs} &= s_r^2 + n s_i^2 + K s_b^2 + n K b^2 \\ V_{is} &= s_r^2 + K s_b^2 \\ V_i &= s_r^2 + k s_i^2, \end{aligned} \quad (14)$$

where  $K$  is as defined in section 9 and  $b$  is the mean value of  $B_j$ .

If tests show that the random trial term  $T_h$  may be omitted, the analysis must be altered to the extent of pooling  $V_e$  and  $V_{gdf1}$  to obtain a new term which may be called the "deviations from regression" mean square, given by

$$V_{dfr} = \frac{SS_i + SS_{it} + Q - R}{n(k-2)}.$$

The estimation equations must also be revised:

$$\begin{aligned} V_{dfr} &= s_r'^2 \\ V_{gs} &= s_r'^2 + K s_b'^2 + n K b'^2 \\ V_{is} &= s_r'^2 + K s_b'^2 \\ V_i &= s_r'^2 + k s_i'^2, \end{aligned} \quad (15)$$

where  $s_r'^2$ ,  $s_b'^2$ ,  $s_i'^2$ , and  $b'$  are new unbiased estimates. Note that the relation between equations (15) and equations (14) is similar to that between equations (3) and equations (2).

It should be observed that the equation for  $V_i$  has the same form in each of the four sets of estimation equations (2), (3), (14), and (15), namely,  $V_i = s_e^2 + k s_i^2$ . In each case,  $s_e^2$  is the measure of residual or unallocated variation, after the variation due to all assignable sources has been eliminated. From a slightly different standpoint, it is a measure of the failure of the performance to follow some law specified by an equation such as

$$X_{jh} = M + I_j + T_h + t_h B_j,$$

or some simpler law. There is some justification for accepting  $s_e^2$  in the more complicated cases represented by equations (2), (14), and (15) as an estimate of the error term  $V_{ei}$  for the simple case in which

there is no trend. By the same token, the intraclass correlation will, in all cases, be estimated by

$$\hat{r} = \frac{s_i^2}{s_i^2 + s_e^2} = \frac{V_i - V_e}{V_i + (k-1)V_e} = \frac{\hat{F}_i - 1}{\hat{F}_i + k - 1}, \quad (16)$$

where  $\hat{F}_i = V_i/V_e$  and  $V_e$  is the measure of residual variation.

It would, of course, be possible to extend equation (13) by including terms of higher degree in  $t_h$ , such as  $t_h^2$ ,  $t_h^3$ , etc. This would, from the standpoint of reliability, merely have the effect of refining the error term,  $V_e$ , removing from it all variation attributable to sources which act through all the individuals of the group. It is to be expected that the reliability would still be estimated by equation (16); the algebraic verification of this possibility is rather complicated, however, and has not yet been carried out.

The use of formula (16) may be illustrated by means of the example of section 9. If we use the simple analysis of variance we obtain  $F_i = 497.90/47.66 = 10.45$  and from formula (11),

$$r^* = 9.45/12.45 = .759.$$

On the other hand, using  $V_e = 28.295$ , the error term from the analysis with the linear component, we have  $\hat{F}_i = 497.90/28.295 = 17.59$ , and from (16),

$$\hat{r} = 16.59/19.59 = .847.$$

In the first case, the error term included a component of variance which could legitimately be separated from it, with the result that we had overestimated the error and thus underestimated the reliability.

#### 11. $r^*$ as average correlation

The close relation of the quantity  $r^*$ , the intraclass correlation adjusted for trend, with the product-moment coefficient of correlation is brought out by the following theorem:

*Theorem.* If, in a table of  $n$  individuals by  $k$  trials the variance of each trial is the same, say  $s^2$ , then the average of the  $k(k-1)/2$  values of the product-moment correlation, calculated for every pair

of trials, is given by  $r^* = \frac{V_i - V_{it}}{V_i + (k-1)V_{it}}$ .

*Proof.* Put  $d_{pq} = X_{jp} - X_{jq}$ , where  $p$  and  $q$  designate any two columns of the table. Then the estimated variance of  $d_{pq}$  is

$$s_{pq}^2 = \frac{1}{n(n-1)} [n \sum d_{pq}^2 - (\sum d_{pq})^2].$$

It will be convenient to employ the following notation:

$$\text{Column totals: } U_h = \sum_j X_{jh}$$

$$\text{Row totals: } W_j = \sum_h X_{jh}$$

$$\text{Grand total: } G = \sum_{j,h} X_{jh}.$$

Then we may write

$$s_{pq}^2 = \frac{1}{n(n-1)} [n \sum (X_{jp} - X_{jq})^2 - (U_p - U_q)^2].$$

On the other hand, if  $r_{pq}$  designates the correlation between columns  $p$  and  $q$ , from equation (5) with  $s_1^2 = s_2^2 = s^2$  we obtain

$$r_{pq} = \frac{2s^2 - s_{pq}^2}{2s^2} = 1 - \frac{s_{pq}^2}{2s^2}.$$

Thus the average value of the correlation,  $\bar{r}$ , is given by

$$\bar{r} = 1 - \frac{\bar{s}^2}{2s^2},$$

where  $\bar{s}^2$  is the average of the  $k(k-1)/2$  values of  $s_{pq}^2$  for all pairs of columns:

$$\begin{aligned} \bar{s}^2 &= \frac{2}{k(k-1)} \sum_{p \neq q} s_{pq}^2 \\ &= \frac{2}{nk(n-1)(k-1)} [n \sum_{j, p \neq q} (X_{jp} - X_{jq})^2 - \sum_{p \neq q} (U_p - U_q)^2]. \end{aligned}$$

In order to simplify this expression, we use the identities

$$\sum (X_{jp} - X_{jq})^2 = k \sum_{j,h} X_{jh}^2 - \sum_j W_j^2,$$

$$\sum (U_p - U_q)^2 = k \sum_h U_h^2 - G^2.$$

Since the proofs of these identities are quite analogous, we prove only the second:

$$\begin{aligned} \sum_{p \neq q} (U_p - U_q)^2 &= (U_1 - U_2)^2 + (U_1 - U_3)^2 + (U_1 - U_4)^2 + \dots \\ &= (k-1)(U_1^2 + U_2^2 + U_3^2 + \dots + U_k^2) \\ &\quad - 2U_1U_2 - 2U_1U_3 - 2U_1U_4 - \dots \end{aligned}$$

But

$$\begin{aligned} G^2 &= (U_1 + U_2 + \dots + U_k)^2 \\ &= U_1^2 + U_2^2 + \dots + U_k^2 + 2U_1U_2 + 2U_1U_3 + 2U_1U_4 + \dots \end{aligned}$$

Hence

$$\sum_{p \neq q} (U_p - U_q)^2 = k(U_1^2 + U_2^2 + \dots + U_k^2) - G^2 = k \sum_h U_h^2 - G^2.$$

Using the two identities,

$$\bar{s}^2 = \frac{1}{(n-1)(k-1)} \left[ \sum_{j,h} X_{jh}^2 - \frac{1}{k} \sum_j W_j^2 - \frac{1}{n} \sum_h U_h^2 + \frac{G^2}{nk} \right].$$

But the expression in the square parenthesis is simply the "interaction sum of squares," which is given by  $(n-1)(k-1)V_{it}$ . Thus

$$\bar{s}^2 = \frac{2}{(n-1)(k-1)} \cdot (n-1)(k-1)V_{it} = 2V_{it}$$

and

$$\bar{r} = 1 - \frac{\bar{s}^2}{2s^2} = 1 - \frac{V_{it}}{s^2}.$$

On the other hand,  $s^2$  is the "within trials" mean square, which may be expressed

$$s^2 = V_{wt} = \frac{1}{k} [V_i + (k-1)V_{it}].$$

Hence

$$\bar{r} = 1 - \frac{kV_{it}}{V_i + (k-1)V_{it}} = \frac{V_i - V_{it}}{V_i + (k-1)V_{it}} = r^*.$$

## 12. Maximum likelihood estimates: $k$ trials

In this section we shall indicate in brief outline how the results of Jackson mentioned in section 7 may be generalized to the case of  $k$  trials, where  $k > 2$ . We require two assumptions in addition to those of section 4:

- (a) The population value,  $\sigma^2$ , of the variance on a particular trial is constant over trials.
- (b) The population value,  $\rho$ , of the correlation between any pair of trials is constant for all pairs of trials.

With these assumptions, the probability density function of the set of  $nk$  variables  $X_{jh}$  is

$$p = K \exp \left[ \frac{B\rho - A(1 - (k-2)\rho)}{2\sigma^2(1-\rho)(1 + (k-1)\rho)} \right], \quad (17)$$

where

$$K = [(2\pi)^k \sigma^{2k} (1-\rho)^{k-1} (1 + (k-1)\rho)]^{-n/2},$$

$$A = \sum_{j,h} (X_{jh} - M_h)^2, \quad B = 2 \sum_{j,p \neq q} (X_{jp} - M_p)(X_{jq} - M_q),$$

and  $M_h$  = the population mean for the  $h^{\text{th}}$  trial.

The maximum likelihood estimates of  $M_p$ ,  $M_q$ ,  $\rho$ , and  $\sigma$  are obtained by differentiating the function  $p$  with respect to these four variables, then setting the derivatives equal to zero, and solving for the quantities in question. The procedure is completely analogous to that exhibited in Jackson (6, 107-112) and need not be given in detail here. We find the following maximum likelihood estimates:

$$\hat{M}_h = \frac{1}{n} \sum_j X_{jh}, \quad \hat{\sigma}^2 = \frac{\hat{A}}{nk}, \quad \hat{\rho} = \frac{\hat{B}}{\hat{A}(k-1)},$$

where  $\hat{A}$  and  $\hat{B}$  have the same form as above, except that in them each  $M$  has been replaced by the corresponding estimate  $\hat{M}$ . On the other hand, it may be shown algebraically that  $\hat{A}$  and  $\hat{B}$  can be expressed in terms of the variance estimates as follows:

$$\hat{A} = k(n-1)(s_i^2 + s_e^2), \quad \hat{B} = k(k-1)(n-1)s_i^2,$$

where  $s_i^2$  and  $s_e^2$  have the same meaning as in section 4. Thus we have

$$\hat{\rho} = \frac{\hat{B}}{\hat{A}(k-1)} = \frac{s_i^2}{s_i^2 + s_e^2} = r^*,$$

where  $r^*$  is the intraclass correlation adjusted for trend as defined in section 5 (d).

If to the assumptions (a) and (b) above we add a third, namely:

(c) The population values of the trial means are all equal to  $M$ ,

then the probability density function is altered merely to the extent of substituting  $M$  for  $M_h$ ,  $M_p$ , and  $M_q$ . Proceeding as before, we find the maximum likelihood estimates

$$\hat{M} = \frac{1}{nk} \sum_{j,h} X_{jh}, \quad \hat{\sigma}^2 = \frac{\hat{A}}{nk}, \quad \hat{\rho} = \frac{\hat{B}}{\hat{A}(k-1)}.$$

where now  $\hat{A}$  and  $\hat{B}$  have the values

$$\hat{A} = (n-1)V_i + n(k-1)V_{wi}, \quad \hat{B} = (k-1)[(n-1)V_i - nV_{wi}],$$

so that

$$\frac{A}{P} = \frac{(n-1)V_i - nV_{wi}}{(n-1)V_i + n(k-1)V_{wi}}.$$

This is the expression which has been referred to as the symmetrical table estimate of the intraclass correlation, section 5(b). It is clearly preferable to use the corresponding unbiased estimate.

### 13. Program for the estimation of reliability

We may sum up the essential results of the preceding developments in the following outline of the procedure for the estimation of reliability when several trials are available:

(a) Test for trend. We need to test for linear as well as for random trend, and consequently the general test of section 9 must be used.

(b) If no trend of any kind is present, then the unbiased estimate of the intraclass correlation is the appropriate estimate [section 5(c)].

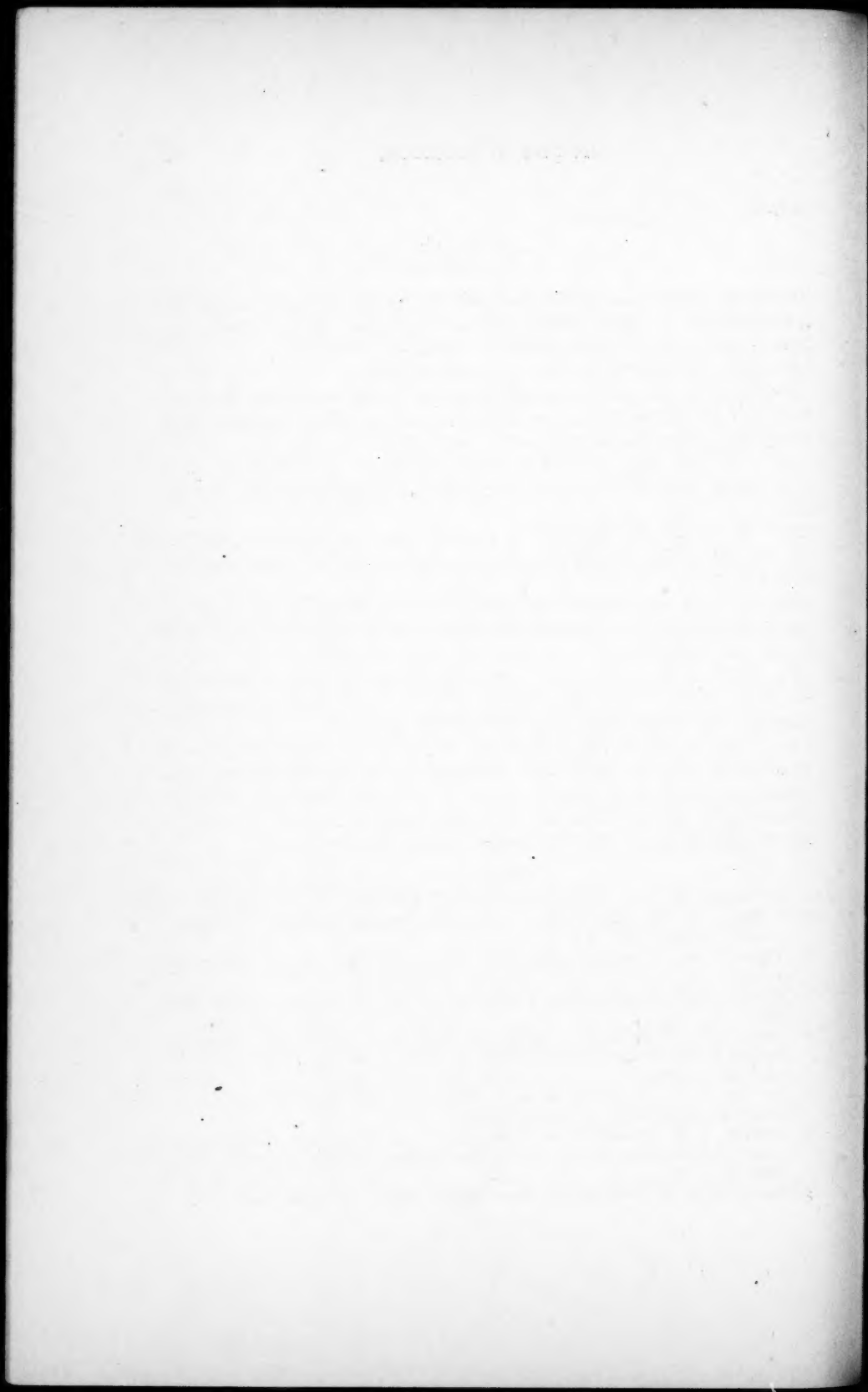
(c) If there is a random fluctuation but no linear trend, then the intraclass correlation adjusted for trend [section 5(d)] should be used to estimate reliability.

(d) If linear trend is present, then the reliability may be estimated by the use of equation (16), where  $V_e$  is the error term in an analysis including a linear term, as outlined in section 9.

(e) Any one of the above methods of estimation may be seriously in error if there is significant heterogeneity of any of the variance estimates, either from trial to trial or from individual to individual. Homogeneity of variance may be investigated by means of tests devised by Bartlett (8, 249-251) and by Welch (5, 40-41).

### REFERENCES

1. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, 43, 533-557.
2. Crump, S. L. The estimation of variance components in analysis of variance. *Biometrics Bull.*, 1946, 2, 7-11.
3. Fisher, R. A. Statistical methods for research workers. London: Oliver and Boyd, 1941.
4. Jackson, R. W. B. Reliability of mental tests. *Brit. J. Psychol.*, 1939, 29, 267-287.
5. Jackson, R. W. B. Application of the analysis of variance and covariance method to educational problems. Bull. No. 11, Dept. of Educ. Research, Univ. of Toronto, 1940.
6. Jackson, R. W. B. Studies on the reliability of tests. Bull. No. 12, Dept. of Educ. Research, Univ. of Toronto, 1941.
7. Jellinek, E. M. Measurements of the consistency of fasting oxygen consumption rates in schizophrenic patients and normal controls. *Biometric Bull.*, 1936, 1, 15-43.
8. Snedecor, G. W. Statistical methods. Ames, Iowa: Collegiate Press, 1946.



## AN EMPIRICAL COMPARISON OF METHODS OF COMMUNALITY ESTIMATION

FRANCIS F. MEDLAND\*  
THE UNIVERSITY OF CHICAGO

This paper shows the formulation of nine methods of estimating the unknown communalities. Each of these methods has been used on experimental data and the results tabulated for comparison. The results show that the most accurate approximations are obtained from the *Centroid No. 1* and the *Graphical* methods.

Communality is defined as that part of the total variance of a test which is due to the common factors. Thus, it differs from the reliability, which is that part of the total variance that is due to the specific factor as well as the common factors. Hence the communality is always less than the reliability unless a specific factor is absent. In such a case, the values are identical.

It is customary to record the reliabilities in the diagonal cells of a table of intercorrelations. In factor analysis, since the interest is centered in the common factor space, it is necessary to use the communalities as the diagonal entries. Thus it is incorrect to record the reliabilities in the diagonal cells unless the variables have been so chosen that they contain no specific factors. However, since communalities are unknown values to be finally determined by factor analysis, they must be estimated.

Fortunately, the estimates of the communalities need not be exact when the number of tests is large. When the number of tests or variables is as small as 10 or 12, it becomes essential to ascertain these values with some degree of exactness. In the case of a small number of variables, it is customary to determine the communalities by factoring the correlation matrix. These communality values are then inserted in the diagonal cells and the matrix is factored again. When the communalities have been determined in this manner, the diagonal elements are not re-estimated for each successive residual table. The purpose of this procedure is to minimize the effect of error variance on the resulting factor matrix. On large batteries, the com-

\* The author wishes to express his appreciation to Professor L. L. Thurstone for his advice and for providing the facilities of the Psychometric Laboratory for this investigation.

putational labor involved makes this method prohibitive. Thus it has become common practice to insert the highest correlation coefficient in a particular column as the diagonal element in that column. When this is done, a new set of diagonal elements is determined for each residual table.

Recently, Holzinger (2) and Thurstone (3) have developed factoring methods that eliminate the necessity of computing a residual table after each factor has been extracted. For the application of these methods, accurate estimation of the communalities is required. With these applications in mind, several methods have been developed (1, 5).

The present study is an analysis of the comparative efficiency of the various methods of estimation. For purposes of the comparison, a sixty-three variable correlation matrix has been used (4). Since the factor matrix for this battery is available, the communalities that have been determined by factoring are known. These values are used as the criterion. Thus for each method here developed, the sixty-three estimated communalities have been obtained. The results and an analysis of the relative accuracy of the various methods are presented.

#### *Methods of Estimation*

##### *a. Graphical Method*

In two nearly collinear tests (1 and 2) the correlations of their respective columns should be proportional to the ratio  $\frac{h_1}{h_2}$  or

$$\frac{r_{2j}}{r_{1j}} = \frac{h_2 h_j \cos \phi_{2j}}{h_1 h_j \cos \phi_{1j}}, \quad (1)$$

and, since the cosines are nearly the same,

$$\frac{r_{2j}}{r_{1j}} = \frac{h_2}{h_1}. \quad (2)$$

With test 1 and test 2 (2 is the test which correlates highest with test 1) as coordinates, the correlations of the other variables with test 1 are plotted against their corresponding correlations with test 2. This should give a linear plot, so that a line, of slope  $s$ , can be passed through these points and the origin by inspection. The slope,  $s_{21}$ , of this line is the estimated ratio,  $\frac{h_2}{h_1}$ , so that

$$r_{2j} = r_{1j} \frac{h_2}{h_1} = r_{1j} s_{21}, \quad (3)$$

and since  $r_{12}$  should be nearly equal to  $h_1 h_2$  it becomes:

$$h_1^2 = \frac{r_{12}}{s_{21}} \quad \text{and} \quad h_2^2 = r_{12} s_{21}, \quad (4)$$

by which the communalities can be estimated.

In the six methods next considered, the solution is determined by the use of sub-groups chosen from the correlation table. In choosing sub-groups, those variables are selected which correlate highest with that variable whose communality is to be determined. Usually three to five variables are chosen. If any coefficient in the sub-group is disproportionately low, that variable is eliminated from the sub-group.

b. Sub-Group Formulas

1. *Centroid No. 1.* The communality to be determined is denoted  $h_1^2$ . Tests 2, 3, ...,  $s$ , are the other variables in the sub-group.

As a first estimate, the highest coefficient in each column of the sub-group is recorded. The sum of the first column, then, is  $(\Sigma r_1 + t_1)$ , where  $\Sigma r_1$  is the sum of all of the known coefficients in the first column and  $t_1$  is the highest coefficient in that column. The sum of all of the coefficients in the sub-group is  $(\Sigma r + \Sigma t)$ , where  $\Sigma r$  is the sum of all the known coefficients and  $\Sigma t$  is the sum of all of the diagonal estimates.

Then:

$$h_1 = \frac{\Sigma r_1 + t_1}{\sqrt{\Sigma r + \Sigma t}} \quad (5)$$

or

$$h_1^2 = \frac{(\Sigma r_1 + t_1)^2}{\Sigma r + \Sigma t}. \quad (6)$$

2. *Centroid No. 2.* The communality to be determined is denoted by  $h_1^2$  and tests 2, 3, ...,  $s$ , are the other variables in the sub-group. The communality of test 1 is estimated by the test projection on the centroid axis of the sub-group. This will, in general, be an underestimation unless the test vector should coincide with the centroid axis. To compensate for this underestimation, the communalities of the adjacent tests are underestimated by entering the average of the adjacent columns in their diagonals. The sum of adjacent

columns 2, 3, ...,  $s$ , becomes  $\frac{s(\Sigma r - \Sigma r_1)}{s-1}$ , where  $\Sigma r$  is the sum of

all of the known coefficients in the table and  $\Sigma r_1$  is the sum of all of the known coefficients in the first column.

The sum of all of the coefficients in the table is estimated as

$$(\sum r_1 + h_1^2) + \frac{s(\sum r - \sum r_1)}{(s-1)},$$

where the complete sum for the first column becomes

$$\sum r_1 + h_1^2.$$

The projection of this test vector on the centroid axis is the ratio of the first column sum to the square root of the sum of all of the coefficients in the table. This projection should be an estimate of  $h_1^2$ .

Then

$$h_1^2 = \frac{(\sum r_1 + h_1^2)^2}{\sum r_1 + h_1^2 + \frac{s(\sum r - \sum r_1)}{(s-1)}}, \quad (7)$$

which simplifies to:

$$h_1^2 = \frac{(s-1)(\sum r_1)^2}{s\sum r - (2s-1)\sum r_1}. \quad (8)$$

For routine computing an equivalent form is developed:

$$h_1^2 = \frac{(s-1)(\sum r_1)^2}{\sum r_1 + 2s\sum r_0}, \quad (9)$$

where  $s$  is the number of tests in the sub-group,  $\sum r_1$  is the sum of all of the known coefficients in column 1, and  $\sum r_0$  is the sum of all of the adjacent coefficients below the diagonal.

3. *Three-Test Formula.* If a cluster of three highly correlated tests is chosen, these test vectors will be separated by very small angles, so that

$$h_1^2 = \frac{r_{1j} r_{1k}}{r_{jk}}. \quad (10)$$

4. *Four-Test Formula.* Expanding the equation (10) to the case of four correlated tests,

$$h_1^2 = \frac{r_{1j} r_{1k}}{r_{jk}} = \frac{r_{1k} r_{1l}}{r_{kl}} = \frac{r_{1j} r_{1l}}{r_{jl}}.$$

Combining these terms and using the geometric mean as an approximation, it becomes:

$$h_1^2 = \sqrt[3]{\frac{r_{1j}^2 r_{1k}^2 r_{1l}^2}{r_{jk} r_{jl} r_{kl}}}. \quad (11)$$

5. *Summation Formula.* Let test 1 be the variable the communality of which is to be determined, and let test 2 be that variable which correlates highest with it. If these two tests have small angular separations, their intercorrelations should be nearly proportional, so that

$$\frac{h_1^2}{r_{12}} = \frac{\sum r_1 - r_{12}}{\sum r_2 - r_{12}}, \quad (12)$$

where  $\sum r_1$  and  $\sum r_2$  are the sums of the known coefficients in columns 1 and 2. From this the computing formula is derived:

$$h_1^2 = \frac{r_{12}(\sum r_1 - r_{12})}{(\sum r_2 - r_{12})}. \quad (13)$$

6. *Spearman Formula.* A generalization of equation (13) can be made for a group of four correlated tests which may be assumed to be nearly collinear. From equation (13), it becomes for the four tests 1, 2, 3, 4,

$$h_1^2 = \frac{r_{12}(\sum r_1 - r_{12})}{\sum r_2 - r_{12}} = \frac{r_{13}(\sum r_1 - r_{13})}{\sum r_3 - r_{13}} = \frac{r_{14}(\sum r_1 - r_{14})}{\sum r_4 - r_{14}}, \quad (14)$$

where  $\sum r_1$ ,  $\sum r_2$ ,  $\sum r_3$ ,  $\sum r_4$ , are the sums of the known coefficients in columns 1, 2, 3, 4, respectively. From equation (14)

$$h_1^2 = \frac{r_{12}(\sum r_1 - r_{12}) + r_{13}(\sum r_1 - r_{13}) + r_{14}(\sum r_1 - r_{14})}{(\sum r_2 - r_{12}) + (\sum r_3 - r_{13}) + (\sum r_4 - r_{14})}, \quad (15)$$

which simplifies to the computing form

$$h_1^2 = \frac{(\sum r_1)^2 - (\sum r_1^2)}{\sum r - 2 \sum r_1}, \quad (16)$$

in which  $\sum r$  is the sum of all of the known coefficients in the sub-group, and  $(\sum r_1^2)$  is the sum of the squares of all of the coefficients in the first column of the sub-group.

7. *Highest Coefficient in the Column.* From equation (10) a simplified estimate may be derived:

$$h_1^2 = \frac{r_{1j} r_{1k}}{r_{jk}}.$$

If these coefficients are of the same order of magnitude, the estimated communality then approaches the highest intercorrelation in the column. While this method is generally applicable to large corre-

lation tables, it is, ordinarily, too unstable for use on small correlation matrices.

8. *Carlson's Method.* This procedure (1) has been developed in connection with a simple multiple-factor approximation method.

This graphical method is generalized for use when more than one factor is present in the correlation matrix. Let test 1 be the test for which the communality is to be determined, and test 2 that test which correlates highest with test 1.

On cross-section paper, an ordinate (test 1) is erected with values from .00 to 1.00 and, at an arbitrary distance of one unit, a second ordinate (test 2) is erected. The correlation coefficients of test 1 with each of the other tests are indicated on ordinate 1, and the correlation coefficients of test 2 with each of the other tests are indicated on ordinate 2. Lines connecting the paired values  $r_{1a}r_{2a}$ ,  $r_{1b}r_{2b}$ , ...,  $r_{1n}r_{2n}$ , are extrapolated to the abscissa. Paired values that are unlike in sign or paired values which are lower on the first ordinate than on the second ordinate are ignored.

The distance on the abscissa from test 1, with the distance between ordinates 1 and 2 as the unit, to the point where each of the remaining lines crosses the abscissa is measured. The sum of these distances divided by the number of tests, plus 1, gives the point on the abscissa from which the communality can be determined. From this point a line is drawn through  $r_{12}$  on ordinate 2 and extrapolated to ordinate 1. This intersection on the first ordinate is the communality estimate for test 1.

### Results

By using each of the methods outlined, communalities were estimated for each variable of the sixty-three test battery. A frequency distribution of absolute deviations from criterion values (Table 1) shows that the *Centroid No. 1* and the graphical methods give results that are most consistent with the criterion. "Root mean square" deviations computed on these data show no significant difference between these two methods on the basis of accuracy. However, the amount of labor involved places the more favorable emphasis upon the *Centroid No. 1* method. A distribution of the algebraic deviations from the criterion (Table 2) is used to determine whether a particular method gives consistent overestimation or underestimation of the true values. Inspection of these distributions and their means does not give significantly different results among the various methods. Further comparison of these tables shows that those methods which depend upon a single correlation coefficient give the least

stable results. This is due to the error in the single correlation coefficient upon which these formulas pivot. The methods that are subject to this criticism are *highest coefficient in the column*, *three-test formula*, and *summation formula*.

Two follow-up studies on smaller batteries (21 variables and 15 variables) show the same results with the exception of "highest coefficient in the column," which proved too unstable for use on the smaller tables.

#### REFERENCES

1. Carlson, Hilding B. A simple orthogonal multiple factor approximation procedure. *Psychometrika*, 1945, 10, 283-301.
2. Holzinger, Karl J. A simple method of factor analysis. *Psychometrika*, 1944, 9, 257-262.
3. Thurstone, L. L. A multiple group method of factoring the correlation matrix. *Psychometrika*, 1945, 10, 73-78.
4. Thurstone, L. L. and Thurstone, T. G. Factorial studies of intelligence. Psychometric Monograph No. 2. Chicago: The University of Chicago Press, 1941.
5. Thurstone, L. L. Multiple factor analysis. Chicago: The University of Chicago Press, 1946.

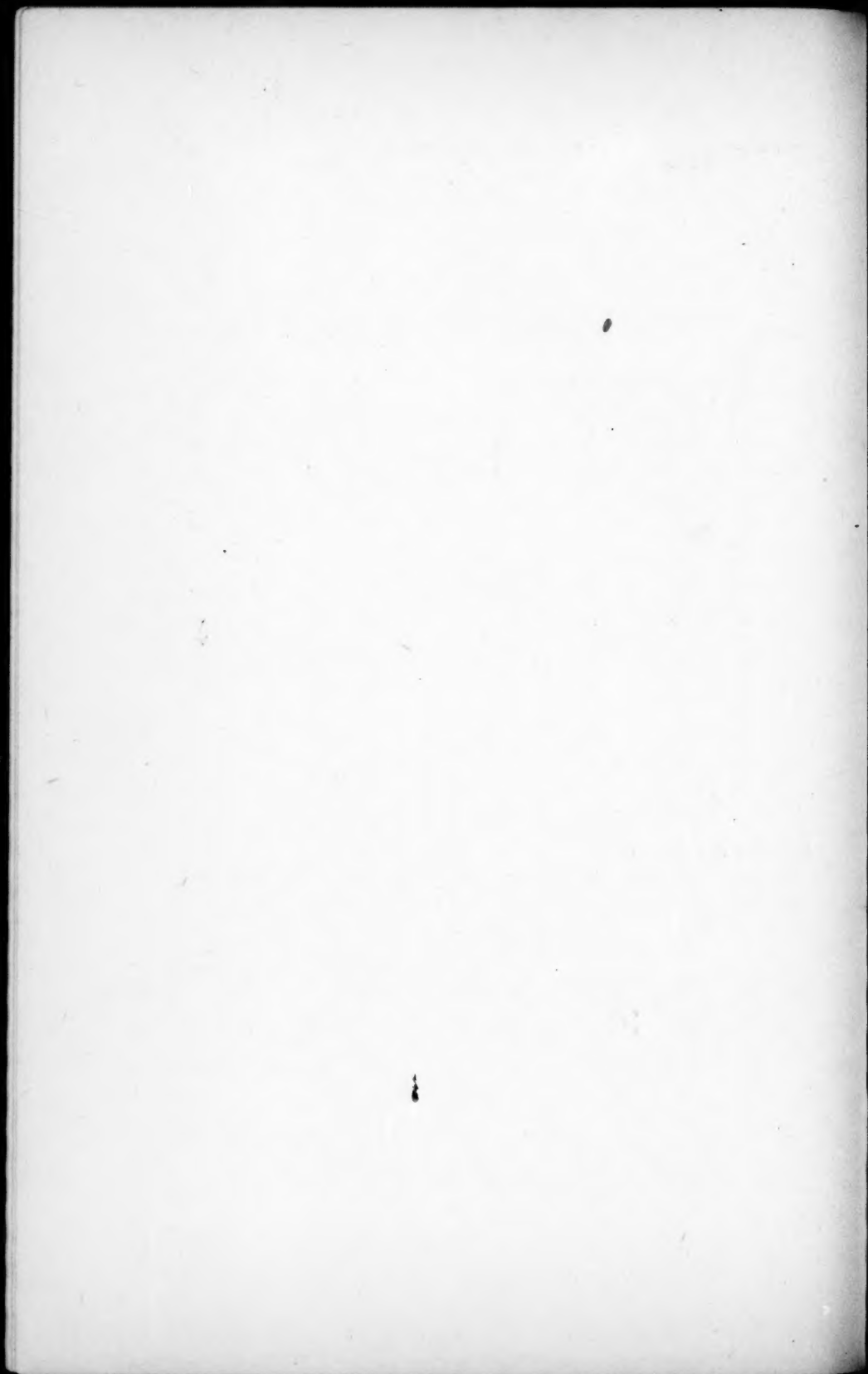
TABLE 1  
Absolute Deviation from Community Obtained by Factoring

Absolute Deviation	Graphical Method	Centroid No. 1	Centroid No. 2	Three-Test Formula	Four-Test Formula	Summation Formula	Spearman Formula	Highest Coeffi- ent in Column	Carlson's Method
over .20		1	3	4	4	3	4	1	2
.20									
.19					1				
.18			1		1	2	1		2
.17	1			1					1
.16	1			3	1		1		
.15		1	1	1					
.14	1			2		2		1	2
.13	1	1		3	3	2	4		1
.12					2	3	1	3	2
.11			3	2	1	2	3	3	3
.10		1	2	4	1	1	1	3	2
.09	4		4	1	3		3	3	2
.08	1	3	1	1	3	2		2	2
.07	4	1	5	1	2	5	6	7	3
.06	5	4	5	5	10	7	6	5	6
.05	7	8	7	8	3	3	6	3	5
.04	8	9	5	8	2	6	4	6	3
.03	9	10	5	6	9	6	4	10	8
.02	8	6	9	3	9	6	7	6	8
.01	9	12	6	7	4	9	8	8	7
.00	4	6	6	3	4	4	4	2	4
N	63	63	63	63	63	63	63	63	63
Root mean square deviation:	.058	.057	.079	.105	.095	.088	.097	.068	.081
Number of deviations $\geq .10$	4	4	10	20	14	15	15	11	15

TABLE 2

Algebraic Deviation from Commuality Obtained by Factoring

Algebraic Deviation	Graphical Method	Centroid No. 1	Centroid No. 2	Three-Test Formula	Four-Test Formula	Summation Formula	Spearman Formula	Highest Coefficient in the Column	Carlson's Method
over .19			2	4	3	3	3		2
.17 .19					1				
.14 .16				4		1		1	2
.11 .13			3	3	4	6	6	5	2
.08 .10	1	3	5	3	6		3	6	2
.05 .07	3	7	8	6	4	9	5	11	3
.02 .04	9	14	10	5	10	8	9	12	9
-.01 .01	13	18	12	10	8	13	12	10	11
-.04 -.02	16	11	9	12	10	10	6	10	9
-.07 -.05	13	6	9	8	11	6	13	4	11
-.10 -.08	4	1	2	3	1	3	1	2	5
-.13 -.11	1	1		2	2	1	2	1	4
-.16 -.14	2	1	1	2	1	1	1		
-.19 -.17	1		1	1	1	2	1		3
under -.19		1	1		1		1	1	
N	63	63	63	63	63	63	63	63	63
Mean deviation from .00:	-.0267	-.0033	.0076	.0162	.0095	.0129	.0081	.0200	-.0128



TABLES FOR THE DETERMINATION OF THE SIGNIFICANCE  
OF SKEWNESS AND OF THE SIGNIFICANCE OF THE  
DIFFERENCE IN THE SKEWNESS OF TWO  
INDEPENDENT DISTRIBUTIONS

ERWIN K. TAYLOR\*

PERSONNEL RESEARCH SECTION, AGO.

Several of the more popular statistics texts used in psychology are reviewed with respect to their treatment of the measurement and interpretation of skewness. Some areas in psychology where measures of skewness of distributions may yield significant information are indicated. Tables of 1%, 2% and 5% points of moment measures of skewness and tables of the 1% and 5% points of the difference in skewness of two uncorrelated distributions are presented. These tables are approximations based on the approximate normality of skewness in large samples from normal populations. The limited confidence with which these tables can be used in the absence of exact knowledge of the distribution function of the underlying statistics is indicated.

It is the purpose of this paper (1) to compare several popular statistical texts in their treatment of skewness, (2) to indicate areas in psychology in which measures of skewness may be important and (3) to present tables of the 1%, 2% and 5% points of skewness and of the 1% and 5% points of the difference in the skewness of two uncorrelated distributions.

It is an almost universal practice in psychological studies involving frequency distributions to present the mean and standard deviation of each such distribution. It is almost as universal to omit measures of symmetry. This omission is a reflection of (or possibly is reflected in) the statistical texts usually used by psychologists for both research and teaching. Confining ourselves for the purpose of this paper to the discussion of measures of symmetry and their use, we might do well to examine several of these texts. Lindquist devotes one paragraph to skewness in his elementary book (13). All that appears in this paragraph is a definition. No method of measuring skewness is given. In his *Statistical Analysis*, Lindquist (12) makes no mention of the phenomenon insofar as a perusal of the index indicates. Edwards (4)

\* The opinions expressed in this article are those of the author and are not to be interpreted as official War Department policy.

presents a section titled "Skewed Distributions" in which he defines both skewness and kurtosis. His last sentence reads, "These are measures of skewness and kurtosis, but we shall have little need for them and they are not included here."

Peters and Van Voorhis (16) give  $\frac{3(M - Mdn)}{\sigma}$  as one of the best measures of skewness.  $\beta_1$  is mentioned. No mention is made of the standard error of  $\frac{3(M - Mdn)}{\sigma}$ . No formula for the standard error of  $\beta_1$  is given; reference, however, is made to sources where it may be found. Aside from the statement that  $\beta_1$  is equal to zero for a normal distribution, nothing further is said of skewness or its implications.

Garrett (8) says, "It is often important to know (1) whether the skewness which so often occurs in distributions of test scores and other measures represents a real divergence from the ideal normal curve; or (2) whether it is the result of chance fluctuation, arising from temporary causes, and is not significant of real disagreement." Why it is important to know whether the skewness is significant or not, the author does not say. Nor does he tell his reader what interpretations are to be made in the event that there is a significant departure from symmetry. He recommends  $\frac{(P_{90} + P_{10})}{2} - 50$  as a measure of skewness and gives its standard error as  $\frac{.5185(P_{90} - P_{10})}{\sqrt{N}}$ , assuming normal distribution of this function.

Guilford in both *Psychometric Methods* (9) and in *Fundamental Statistics* (10) goes somewhat more fully into the matter. In the former volume (p. 105) he says,

There are a number of possible causes for skewed distributions. One cause is a selective factor, or set of factors, which cause the measurements to deviate from a random sampling. If in testing the intelligence of fourteen-year-old boys we used only those below the ninth grade, we would probably find a distribution that is positively skewed because the brighter fourteen-year-old boys in the high-school grades had been eliminated. Another frequent cause in mental testing is the use of a test that is too difficult or too easy for the group to which it is applied . . . Another cause may be the measuring instrument. If the units of amount near the lower end of the scale are in reality larger than those at the up-

per end, there will be an apparent bunching of cases at the lower end and therefore a positive skewing effect. If we eliminate all these factors causing artificial skewing, and other such factors as well, the skewness may express the actual facts. The distribution may in reality be one lacking in symmetry. Should this be distinctly so, the application of the normal curve is questionable and one of Pearson's other types should be used.

In addition to the importance of skewness in the fields of mental and educational testing pointed out by Guilford, investigators in other areas of psychology have employed departures from symmetry as highly meaningful characteristics of their measures or populations. Katzoff (11) reworking Allport and Solomon's (1) data found skewness a much more satisfactory index of conforming behavior than was the "J-curve" hypothesis proposed in the original research. Both Yoder (20) and Ford (7) indicate the possible influence of restriction of output upon the symmetry of production curves, presenting some data to indicate negative skewness under such circumstances. Bliss (2) found that artificial restriction of output resulted in negatively skewed distributions of production, while poor motivation resulted in positively skewed distributions. The apparent contradiction between the findings of Bliss and those of Yoder and Ford vanishes upon analysis. In the case of the latter writers, the situation was one in which there was high motivation to reach a given point and fear of rate cut if that point was exceeded. This would naturally result in a negatively skewed distribution. Bliss, too, secured such a distribution when the output of his subjects was artificially restricted beyond a certain point by the speed of the machines on which they worked.

The writer in some unpublished research in a sales organization found that the production (sales) distribution shifted from a positively to a negatively skewed curve when high motivation in the form of well conducted sales contests was introduced into the situation. In the case of individual salesmen, those whose distributions of sales before the contest were positively skewed generally showed greater gain than those with similar means but negatively skewed or symmetrical distributions.

An inspection of Uhrbrock's (19) data on the attitude of factory workers, clerks, and foremen reveals the extent to which departures from symmetry are related to attitudinal changes. Another psychological field in which skewness may have important implications is that of learning. While the writer can cite no evidence to support this contention, it seems logical that for any given individual who has

learned a particular skill completely, the distribution of the accuracy or of the speed (where not artificially restricted by zero time) with which repeated trials are accomplished should be symmetric about the mean accuracy or speed of the individual. Motivation in this case should of course be ruled out. The novice on whom the distribution of repeated trails is available, it is hypothesized, would be inclined to present a positively skewed distribution. As learning progressed, the distribution should approach symmetry as a limit. If this relationship could be empirically demonstrated, it might prove quite useful in many learning experiments.

It is apparent that there are several areas in psychology in which the significance of the departure of distributions from symmetry may be of considerable importance. By the same token there are situations in psychological research in which it is desirable to determine whether the skewness of one distribution is greater than that of another. As in the case of all statistics, such determinations must be made in terms of the standard error of the measures employed.

Of the many measures of skewness,  $\alpha_3 = \sqrt{\beta_1}$  or Fisher's\* (6)

$$g_1 = [N^2 / (N-1)(N-2)] \alpha_3 \quad (1)$$

are probably most satisfactory from the point of view of *both* statistical efficiency *and* computational ease. A method for the systematic computation of  $\alpha_3$  is given by Cureton and Dunlap (3).  $\alpha_3$  may of course be easily converted to  $g_1$  as indicated above. The systematic computation of  $g_1$  is given by Snedecor (17).

Having determined the value of skewness of a given distribution, it becomes necessary to determine whether or not such skewness is significantly different from zero. Unfortunately, the distribution function of neither of the recommended measures of skewness has been determined. The basic work of Fisher (5) has established the variance of  $g_1$  (and necessarily of  $\alpha_3$ ) from normally distributed populations as

\* This is predicated on the computation of  $\sigma$  by the formula

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{(N-1)}}$$

If

$$\sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

is used in the computation of  $\sigma$ , the relationship between  $\alpha_3$  and  $g_1$  becomes

$$g_1 = \left( \frac{N-1}{N} \right)^{3/4} \left( \frac{N^2}{(N-1)(N-2)} \right) \alpha_3.$$

$$\sigma_{g_1}^2 = \frac{6N(N-1)}{(N-2)(N+1)(N+3)}. \quad (2)$$

Theoretical as well as empirical evidence tends to indicate that the distribution of  $g_1$  (or  $\alpha_3$ ) is approximately normal for samples drawn from a normal population. Pepper (15) drew 700 samples of size 10 from a normal population. The distribution of the measures of  $\alpha_3$  was tested for departure from normality by chi squared. It was found that departures from normality were less than would be found by chance alone 97% of the time. So good a fit indicates the feasibility of using normal curve area tables to test the probability of an obtained  $\alpha_3$  (or  $g_1$ ) having arisen in a sample from a symmetrical universe.

Pearson (14), on the basis of the work of Fisher (5), published tables of the 1% and 5% points of  $\alpha_3$  for samples of size 50 to 5000. These were extended by Williams (18) to samples of size 25. These tables are based on the formula

$$\sigma_{\alpha_3} = \sqrt{6/N}.$$

Table 1 presents the 1%, 2%, and 5% points of both  $\alpha_3$  and  $g_1$  on samples of the same range of  $N$  but using the more exact formula

$$\sigma_{g_1} = \left[ \frac{6N(N-1)}{(N-2)(N+1)(N+3)} \right]^{1/2}. \quad (3)$$

In addition, Table 1 gives the variance and standard error of  $g_1$  as well as the constants by means of which  $\alpha_3$  may be converted to  $g_1$  and  $g_1$  to  $\alpha_3$ . This table may be used to determine the significance of departure of skewness from zero at any of three levels.

Table 2 presents the 1% points of the difference in skewness (as measured by  $g_1$ ) of two uncorrelated distributions of size from  $N = 30$  to  $N = 5000$  for either. Table 3 presents the 5% points of the same statistics. In using these tables it should be noted that these points are based on the formula

$$\sqrt{\sigma_{g_{1a}}^2 + \sigma_{g_{1b}}^2} \cdot K, \quad (4)$$

where

$K = 2.58$  for Table 2 and

$K = 1.96$  for Table 3.

This neglects the  $(2r \sigma_{g_a} \sigma_{g_b})$  term common to formulas for the standard error of the difference between two statistics. These tables should be used only in connection with uncorrelated variables.

Where it is required to determine the significance of the difference

between correlated skewness measures and that correlation is known, the proper formula may be easily derived from the  $\sigma^2_{g_1}$  or  $\sigma^2_{g_2}$  columns of Table 1. The needed statistic is

$$\sigma_{diff} = \sqrt{\sigma^2_{g_{1a}} + \sigma^2_{g_{1b}} - 2r_{g_{1a}g_{1b}} \sigma_{g_{1a}} \sigma_{g_{1b}}}. \quad (5)$$

In attempting to determine the significance of the difference in skewness of two samples drawn from non-normal populations, the writer has been unable to find any evidence with respect to the distribution of such differences. Undoubtedly considerable research into the question is required before Tables 2 and 3 can be used with complete confidence. The assumption of normality without proof thereof is not without considerable precedent in the field of statistics. In general, it appears that the statistics from non-normal universes are more nearly normal than are their parent populations and that this tendency toward the Gaussian curve is a function of  $N$ . Until the distribution functions of the statistics under consideration have been developed, the tables presented can probably be used in practice without doing extensive violence to the data.

TABLE 1  
Standard Errors and Fiducial Limits of Skewness

(1) $N$	(2) $\sigma^2 g_1$	(3) $\sigma g_1$	(4) $\frac{N^2}{(N-1)(N-2)}$	(5) $\frac{(N-1)(N-3)}{N^2}$	(6) 5% L. $g_1$
5000	.001199	.034626	1.00060	.999400	.067866
4000	.001499	.038716	1.00075	.999250	.075882
3000	.001998	.044698	1.00100	.999000	.087606
2000	.002995	.054726	1.00150	.998500	.107261
1500	.003992	.063182	1.00200	.998001	.123824
1000	.005982	.077343	1.00301	.997002	.151589
900	.006644	.081511	1.00334	.996669	.159758
800	.007472	.086441	1.00376	.996253	.169421
700	.008534	.092379	1.00430	.995718	.181059
600	.009950	.099749	1.00502	.995006	.195504
500	.011928	.109215	1.00603	.994008	.214057
450	.013245	.115086	1.00670	.993343	.225564
400	.014888	.122016	1.00754	.992512	.239146
350	.016997	.130369	1.00863	.991445	.255518
300	.019802	.140719	1.01008	.990022	.275804
250	.023716	.154000	1.01211	.988032	.301843
200	.029558	.171924	1.01518	.985050	.336964
175	.033711	.183605	1.01737	.982922	.359858
150	.032919	.198037	1.02032	.980089	.388145
125	.046881	.216520	1.02446	.976128	.424371
100	.058264	.241379	1.03072	.970200	.473093
90	.064532	.254031	1.03422	.966914	.497891
80	.072312	.268909	1.03862	.962812	.527051
70	.082225	.286749	1.04433	.957551	.562017
60	.095292	.308694	1.05202	.950556	.605028
50	.113300	.336600	1.06293	.940800	.659723
45	.125126	.353731	1.07030	.934321	.693299
40	.139714	.373783	1.07962	.926250	.732600
35	.158160	.397693	1.09180	.915918	.779462
30	.182237	.426892	1.10837	.902222	.872994
25	.215002	.463683	1.13225	.883200	.957042

TABLE 1 (continued)

<i>N</i>	(7) 2% L. $g_1$	(8) 1% L. $g_1$	(9) 5% L. $\alpha_3$	(10) 2% L. $\alpha_3$	(11) 1% L. $\alpha_3$
5000	.080552	.089190	.067825	.080504	.089136
4000	.090067	.099725	.075825	.089999	.099650
3000	.103983	.115134	.087184	.103879	.115019
2000	.127311	.140964	.107100	.127120	.140753
1500	.146983	.162745	.123586	.146689	.162420
1000	.179926	.199222	.151135	.179387	.198625
900	.189622	.209958	.159226	.188990	.209259
800	.201091	.222656	.168786	.200338	.221822
700	.214905	.237952	.180284	.213985	.236933
600	.232050	.256935	.194528	.230891	.255652
500	.254071	.281318	.212774	.252549	.279632
450	.267729	.296441	.224062	.265947	.294468
400	.283851	.314291	.237355	.281726	.311938
350	.303283	.336171	.253332	.300688	.333295
300	.327360	.362467	.273052	.324094	.358850
250	.358256	.396676	.298222	.353968	.391929
200	.399954	.442845	.331926	.393975	.436224
175	.427128	.472933	.353712	.419834	.464856
150	.460701	.510108	.380417	.451528	.499951
125	.503699	.557717	.414240	.491675	.544403
100	.561530	.621749	.458995	.544796	.603221
90	.590962	.654338	.481418	.571409	.632869
80	.625574	.692661	.507451	.602310	.666902
70	.667076	.738614	.538160	.638759	.707261
60	.718127	.795140	.575113	.682620	.755825
50	.783046	.867021	.620667	.736690	.815693
45	.822899	.911147	.647764	.768852	.851304
40	.869546	.962798	.678571	.805602	.891792
35	.925169	1.024386	.713923	.847379	.938254
30	1.051008	1.176514	.787634	.948243	1.061477
25	1.155498	1.296921	.845259	1.020536	1.145441

TABLE 2\*  
1% Limits of  $g_{1x} - g_{1y}$

N	5000	4000	3000	2000	1500	1000	900
5000	1261						
4000	1338	1410					
3000	1456	1523	1628				
2000	1663	1722	1816	1986			
1500	1856	1909	1994	2149	2302		
1000	2183	2228	2301	2437	2572	2817	
900	2281	2324	2395	2526	2656	2894	2969
800	2399	2440	2507	2632	2758	2988	3060
700	2533	2572	2635	2756	2876	3097	3167
600	2720	2756	2816	2928	3041	3251	3318
500	2951	2985	3040	3144	3250	3447	3510
450	3096	3128	3180	3280	3382	3572	3633
400	3267	3297	3347	3442	3539	3721	3780
350	3475	3573	3550	3640	3732	3905	3960
300	3733	3759	3803	3887	3973	4136	4189
250	4066	4090	4130	4208	4288	4439	4488
200	4517	4539	4576	4646	4718	4856	4901
175	4813	4833	4867	4933	5001	5132	5174
150	5178	5198	5229	5291	5354	5476	5516
125	5648	5666	5695	5751	5810	5922	5959
100	6281	6297	6323	6374	6427	6529	6562
90	6604	6619	6644	6692	6743	6840	6872
80	6984	6998	7022	7067	7115	7207	7238
70	7440	7453	7475	7518	7563	7650	7679
60	8001	8014	8034	8074	8116	8197	8224
50	8716	8727	8746	8783	8821	8896	8921
45	9155	9166	9184	9219	9256	9327	9350
40	9669	9679	9697	9730	9765	9832	9854
35	10283	10292	10308	10340	10372	10436	10457
30	11032	11041	11056	11085	11116	11175	11195

\* Entries in Tables 2 and 3 have been multiplied by 104.

TABLE 2 (continued)

N	800	700	600	500	450	400	350	300
5000								
4000								
3000								
2000								
1500								
1000								
900								
800	3148							
700	3253	3355						
600	3400	3496	3634					
500	3588	3679	3810	3978				
450	3707	3796	3923	4087	4192			
400	3852	3937	4060	4218	4320	4445		
350	4029	4110	4228	4381	4479	4599	4749	
300	4254	4331	4443	4588	4683	4789	4941	5126
250	4549	4621	4726	4863	4952	5061	5197	5373
200	4957	5023	5120	5246	5329	5430	5558	5723
175	5227	5290	5382	5503	5582	5678	5800	5959
150	5566	5625	5712	5825	5900	5992	6107	6258
125	6005	6060	6141	6247	6316	6402	6510	6652
100	6604	6654	6727	6824	6888	6967	7066	7197
90	6912	6960	7030	7123	7184	7259	7355	7480
80	7276	7321	7388	7476	7534	7606	7698	7818
70	7714	7757	7820	7904	7959	8027	8114	8228
60	8257	8297	8356	8434	8486	8550	8631	8739
50	8952	8989	9043	9115	9163	9222	9298	9397
45	9380	9415	9467	9536	9582	9638	9711	9806
40	9882	9916	9965	10031	10074	10128	10197	10288
35	10483	10515	10561	10623	10664	10715	10780	10866
30	11219	11249	11292	11350	11389	11436	11497	11578

TABLE 2 (continued)

N	250	200	175	150	125	100	90
250	5610						
200	5945	6263					
175	6173	6479	6688				
150	6462	6755	6956	7214			
125	6844	7122	7312	7558	7887		
100	7375	7633	7812	8042	8352	8793	
90	7652	7901	8074	8297	8598	9026	9254
80	7982	8221	8387	8602	8893	9308	9529
70	8384	8612	8771	8976	9255	9655	9868
60	8886	9101	9252	9447	9712	10094	10298
50	9535	9736	9876	10060	10309	10669	10862
45	9938	10131	10266	10442	10683	11031	11218
40	10413	10598	10727	10896	11127	11461	11641
35	10985	11160	11283	11444	11664	11983	12155
30	11690	11854	11970	12122	12329	12632	12796

TABLE 2 (continued)

N	80	70	60	50	45	40	35	30
250								
200								
175								
150								
125								
100								
90								
80	9796							
70	10126	10446						
60	10545	10853	11245					
50	11097	11390	11764	12262				
45	11445	11729	12093	12577	12886			
40	11861	12135	12487	12957	13256	13616		
35	12366	12629	12968	13420	13710	14058	14487	
30	12996	13246	13570	14003	14280	14615	15028	15551

TABLE 3

5% Limits of  $g_{1x} - g_{1y}$ 

$N$	5000	4000	3000	2000	1500	1000	900
5000	0960						
4000	1018	1073					
3000	1108	1159	1239				
2000	1265	1310	1382	1511			
1500	1412	1453	1517	1635	1752		
1000	1661	1695	1751	1854	1957	2143	
900	1736	1768	1822	1922	2021	2202	2259
800	1825	1857	1908	2003	2099	2274	2328
700	1927	1957	2005	2097	2188	2357	2410
600	2070	2097	2143	2228	2314	2474	2525
500	2245	2271	2313	2392	2473	2623	2671
450	2356	2380	2420	2496	2573	2718	2764
400	2486	2509	2547	2619	2693	2831	2876
350	2644	2719	2701	2770	2840	2971	3013
300	2840	2860	2894	2958	3023	3147	3187
250	3094	3112	3143	3202	3263	3378	3415
200	3437	3454	3482	3535	3590	3695	3729
175	3662	3677	3703	3754	3805	3905	3937
150	3940	3955	3979	4026	4074	4167	4197
125	4298	4311	4333	4376	4421	4506	4534
100	4779	4791	4811	4850	4890	4968	4993
90	5025	5036	5055	5092	5131	5205	5229
80	5314	5325	5343	5377	5414	5484	5507
70	5661	5671	5688	5721	5755	5821	5843
60	6088	6098	6113	6144	6176	6237	6258
50	6632	6640	6655	6683	6712	6769	6788
45	6966	6974	6988	7015	7043	7097	7115
40	7357	7365	7379	7404	7430	7481	7498
35	7824	7831	7843	7868	7892	7941	7957
30	8394	8401	8413	8435	8458	8503	8518

TABLE 3 (continued)

N	800	700	600	500	450	400	350	300
5000								
4000								
3000								
2000								
1500								
1000								
900								
800	2395							
700	2475	2553						
600	2587	2660	2765					
500	2730	2799	2899	3027				
450	2821	2888	2985	3110	3190			
400	2931	2996	3089	3210	3287	3382		
350	3066	3127	3217	3334	3408	3499	3614	
300	3237	3296	3381	3491	3563	3651	3760	3900
250	3461	3516	3596	3700	3768	3851	3954	4088
200	3772	3822	3896	3992	4055	4132	4229	4355
175	3977	4025	4095	4187	4247	4320	4413	4534
150	4235	4280	4346	4432	4489	4559	4647	4762
125	4569	4611	4673	4753	4806	4871	4954	5062
100	5025	5063	5119	5192	5241	5301	5377	5476
90	5259	5296	5349	5420	5466	5523	5596	5692
80	5536	5571	5622	5689	5733	5787	5857	5949
70	5870	5902	5950	6014	6056	6108	6174	6261
60	6283	6313	6358	6418	6457	6506	6567	6650
50	6812	6840	6881	6936	6972	7017	7075	7150
45	7137	7164	7204	7256	7291	7334	7389	7461
40	7519	7545	7582	7633	7665	7706	7759	7828
35	7977	8001	8036	8083	8114	8153	8203	8236
30	8537	8559	8592	8636	8666	8702	8748	8810

TABLE 3 (continued)

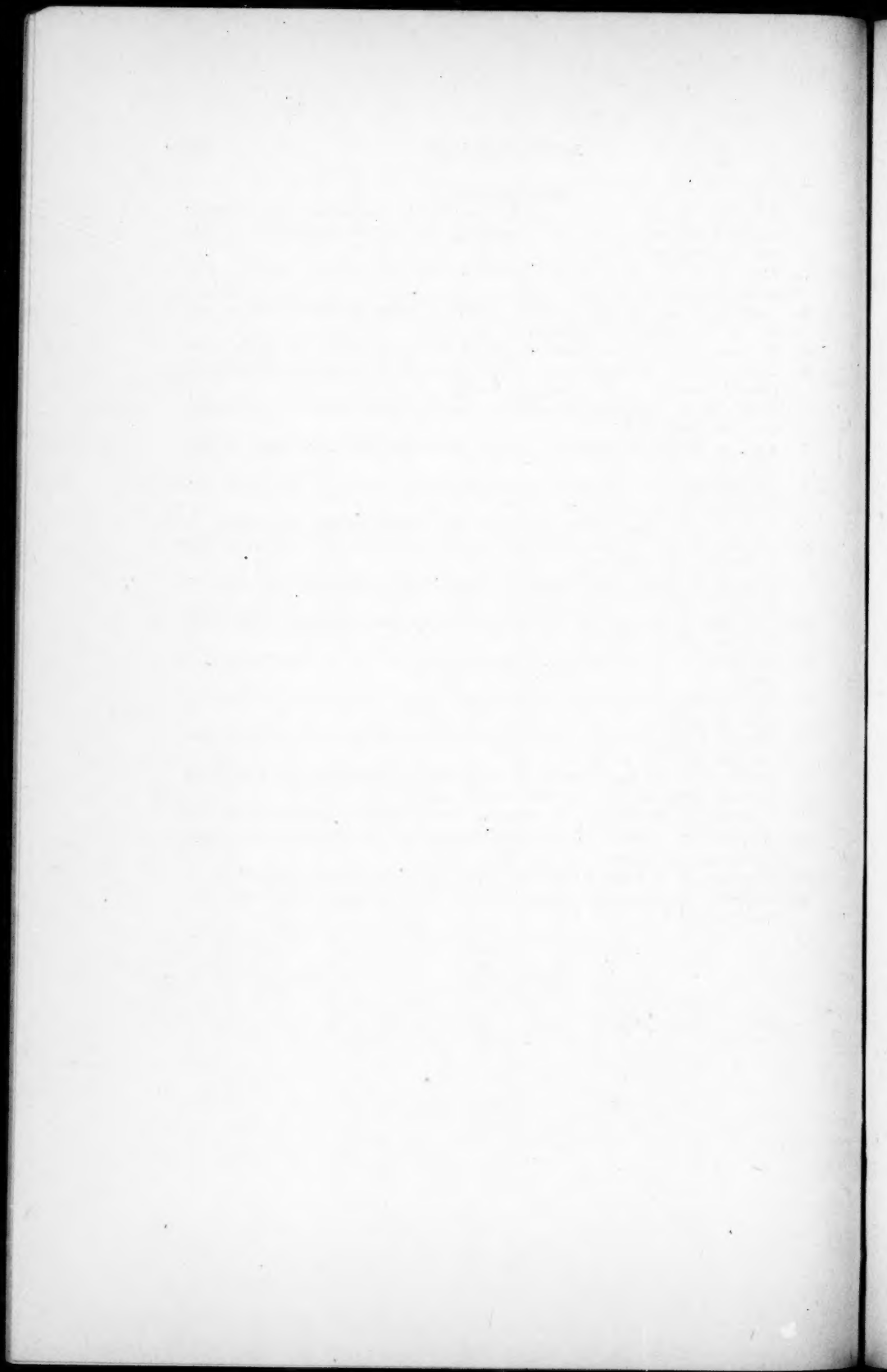
N	250	200	175	150	125	100	90
250	4269						
200	4524	4766					
175	4697	4930	5089				
150	4917	5140	5293	5489			
125	5208	5419	5564	5751	6001		
100	5612	5808	5944	6119	6355	6691	
90	5822	6012	6144	6313	6542	6868	7041
80	6074	6255	6382	6545	6767	7083	7251
70	6379	6553	6674	6830	7042	7347	7509
60	6761	6925	7040	7188	7390	7681	7836
50	7255	7408	7515	7655	7844	8118	8265
45	7562	7709	7811	7945	8129	8394	8536
40	7923	8064	8162	8291	8467	8721	8858
35	8359	8492	8585	8708	8875	9118	9249
30	8895	9020	9108	9224	9381	9612	9737

TABLE 3 (continued)

N	80	70	60	50	45	40	35	30
250								
200								
175								
150								
125								
100								
90								
80	7454							
70	7705	7948						
60	8024	8258	8556					
50	8444	8667	8951	9330				
45	8709	8925	9202	9570	9805			
40	9025	9243	9501	9859	10087	10361		
35	9409	9610	9867	10211	10432	10697	11023	
30	9889	10079	10326	10655	10866	11121	11435	11833

## REFERENCES

1. Allport, F. H. and Solomon, R. S. Lengths of conversation: a conformity situation analyzed by the telec continuum and J-curve hypotheses. *J. abn. soc. Psychol.*, 1939, 34, p. 419.
2. Bliss, E. F., Jr. Earnings of machine tenders and of bench workers. *Personnel J.*, 1931, 10, p. 102.
3. Cureton, E. E. and Dunlap, J. W. Note on testing of departure from normality. *J. appl. Psych.*, 1930, 14, p. 91.
4. Edwards, Allen L. Statistical analysis. New York: Rinehart & Co., 1946.
5. Fisher, R. A. Moments and product moments of sampling distributions. *Proc. London Math. Soc.*, 1929, 50, p. 199.
6. Fisher, R. A. Statistical methods for research workers. (6th Ed.) Edinburg: Oliver & Boyd, 1936.
7. Ford, A. A scientific approach to labor problems. New York: McGraw-Hill, 1931.
8. Garrett, Henry E. Statistics in psychology and education. (2nd Ed.) New York: Longmans Green & Co., 1938.
9. Guilford, J. P. Psychometric methods. New York: McGraw-Hill, 1936.
10. Guilford, J. P. Fundamental statistics in psychology and education. New York: McGraw-Hill, 1942.
11. Katsoff, E. T. The measurement of conformity. *Psychometrika*, 1942, 7, p. 31.
12. Lindquist, E. F. Statistical analysis in educational research. New York: Houghton-Mifflin, 1940.
13. Lindquist, E. F. A first course in statistics. New York: Houghton-Mifflin, 1942.
14. Pearson, E. S. A further development of tests of normality. *Biometrika*, 1931, 22, p. 239.
15. Pepper, J. The sampling distribution of the third moment coefficient. *Biometrika*, 1932, 24, p. 53.
16. Peters, C. C. and Van Voorhis, W. R. Statistical procedures and their mathematical bases. New York: McGraw-Hill, 1940.
17. Snedecor, G. W. Statistical methods. Ames, Iowa: Collegiate Press, 1938.
18. Williams, P. Note on the sampling distribution of  $\sqrt{\beta_1}$ , where the population is normal. *Biometrika*, 1935, 27, p. 269.
19. Uhrbrock, R. S. Attitudes of 4430 employees. *J. soc. Psychol.*, 1934, 5, p. 365.
20. Yoder, D. Personnel management and industrial relations. New York: Prentice Hall, 1942.



## A MODIFIED AITKEN PIVOTAL CONDENSATION METHOD FOR PARTIAL REGRESSION AND MULTIPLE CORRELATION

ALICE VAN BOVEN  
REDLANDS, CALIFORNIA

In this modification of the Aitken Pivotal Condensation Method for obtaining partial regression weights and multiple correlation coefficients, the unit matrices used by Aitken are eliminated, with a resultant decrease in the computational work. Through a simplified arrangement of successive matrices, each one is reduced by the same simple rule. Other modifications also contribute to ease of computation and facilitate checking.

The method of using determinants to solve a system of simultaneous equations has long been accepted. An excellent comprehensive survey of a number of such methods can be found in an article by P. S. Dwyer in *Psychometrika*, 1941, 6, 101-129. A partial regression equation for three or more variables is constructed by solving as many simultaneous equations as there are variables. Each of these equations is derived mathematically from the assumption that the variance of the error of estimate is to be reduced to a minimum when each component in turn is permitted to vary while all the others are held constant.

Godfrey Thomson in his book *The Factorial Analysis of Human Ability*\* describes the Aitken Pivotal Condensation Method for obtaining the regression weights and multiple  $R$  by employing a series of matrices. An example which employs the same general technique and nearly the same steps is worked out in an article by Godfrey Thomson in the *Journal of Educational Psychology*, 1936, 27, 51. Both of these examples of the Aitken method start with a matrix of first-order correlations and a unit matrix beside it. These two matrices are gradually condensed until the beta coefficients are obtained. Two Fisher methods employing unit matrices are shown by Dwyer on p. 122 of the article referred to above.

In the method here described, the unit matrices employed by Aitken are eliminated. The many zeroes in the unit matrices are not needed to obtain the final results. The total number of arithmetical calcu-

\* Thomson, Godfrey. *The factorial analysis of human ability*, Boston: Houghton Mifflin, 1939, pp. 89-95.

lations is thus materially reduced. A further contribution toward compactness and ease of computation is found in the series of matrices set up at the right-hand side of the computation sheet. (These matrices are labeled  $A'$ ,  $B'$ ,  $C'$ , and  $D'$  in the sample problem shown in Table 1.) It adds much to the ease of application to have a space provided on the computation sheet for the figures derived by multiplication which in turn are to be subtracted from numbers in the original matrix. This multiplication must be done anyway when solving determinants or matrices set up for the purpose of solving a series of simultaneous equations. Writing them out in matrix form shows the symmetric pattern obtained and also provides the conventional check by addition at the end of each row, thus decreasing the labor and the chance for errors to slip in at these points. The figures are all on the page, so that when rechecking is necessary, time is saved.

There are in all but the last two matrices ( $C$ ,  $D$ ,  $C'$ , and  $D'$  in the sample) some symmetric numbers, which, as far as the final results are obtained, could be omitted from the computation sheet, exactly as in several of the examples surveyed by Dwyer. For the sake of the check by addition of each row, it is, in the opinion of the author, better to copy them on the computation sheet.

Aitken\* describes the use of a pivot and pivotal condensation for the solution of determinants as follows:

A determinant of order  $n$  being reduced by a first pivotal condensation to one of order  $n-1$ , the latter in its turn can be reduced by a second pivotal condensation to one of order  $n-2$ , and so on until finally we have a determinant of order 2, a simple cross-product. Naturally the case where the pivot is 1 or  $-1$  will give the least arithmetical work. It is always possible, by a preliminary division of the row or column containing any non-zero element, to turn that element into a unit pivot.

The chief advantage, for practical purposes, in the method here presented is in the simplicity of arrangement, in that each matrix is condensed by the same simple rule which is readily learned.

A description of a sample application of the modification of Aitken's method is presented step by step so that it can be followed by comparing the text with the sample problem appended (Table 1).

Let us say there are  $n$  independent variables, labeled  $V_1$ ,  $V_2$ ,  $V_3$ , ...,  $V_n$ . There is also the criterion,  $C$ . The method consists of  $n$

\* Aitken, A. C. *Determinants and matrices*, Edinburgh: Oliver and Boyd, 1939, p. 47.

double matrices, arranged successively and solved in the same manner, so that once the process is learned, the work is largely routine.

When laying out the problem on paper, allowance should be made for  $2n + 2$  columns across the page. It is a great help to enter all negative values in red throughout the problem.

*Matrix A.* This is simply a correlation matrix of the  $n$  independent variables, with the criterion correlations added on a row below and two columns added at the right. The rows and columns of the symmetrical square wherein are entered correlation values of the  $n$  independent variables are labeled  $V_1, V_2, V_3, \dots, V_n$ , respectively, and will henceforth be so designated. The bottom row of the matrix in which are entered correlation values between the independent variables and the criterion will be designated as  $C$ . To the right of the  $V_n$  column is the  $-1$  column, where  $-1$  is entered in all the  $V$  series rows, but  $0$  is placed in the  $C$  row. The last column of Matrix  $A$  is the total column, designated as  $T$ . In it are entered the algebraic sums of each row of the matrix. Note that in the  $V$  series rows the  $1$  and  $-1$  will cancel each other in adding, but there is neither a  $1$  nor a  $-1$  in row  $C$ .

*Matrix A'.* The columns in this matrix are labeled  $V_2, V_3, \dots, V_n$ , and  $T$ . In the top row of this matrix the numbers .481, .146, .118, and .745 are copied from Matrix  $A$ . Note that the .745 copied into column  $T$  is the algebraic sum of the other three copied numbers.

The symmetrical part of Matrix  $A'$  is in reality a submatrix containing  $n - 1$  arrays. In the example, it includes the nine numbers in columns  $V_2, V_3$ , and  $V_n$  and in the rows  $V_2, V_3$ , and  $V_n$ . These numbers and the numbers in the  $T$  column are obtained by multiplying each number in the top row of Matrix  $A'$  by the numbers in column  $V_1$  of Matrix  $A$ . In the sample problem, when .481 from row  $V_2$ , column  $V_1$  is multiplied by the numbers .481, .146, .118, and .745 from the top row of Matrix  $A'$  the results are, respectively, .231, .070, .057, and .358. This row is checked by seeing that the number in the  $T$  column equals the algebraic sum of the preceding numbers. When .146 is used as multiplier for the numbers in the top row of Matrix  $A'$ , the results are .070, .021, .017, and .109. The last number equals the algebraic sum of the other three. Likewise when .118 is used as multiplier with the same four numbers in the top row of Matrix  $A'$ , the results are .057, .017, .014, and .088. The last named value, .088, equals the sum of the other three. The .490 in row  $C$ , column  $V_1$  must also be multiplied by each number in the top row of Matrix  $A'$ . The results in this case are .236, .072, .058, and .365.

A single setting of the slide rule is sufficient for the multiplication in each row. Each row of each matrix should be checked by ad-

TABLE 1  
Sample Problem in Multiple Correlation  
Job Information Test in Military Correspondence Versus Four Other Variables  
(94 cases, all enlisted men)

	$V_1$	$V_2$	$V_3$	$V_n$	$T$	$V'_2$	$V'_3$	$V'_n$	$T'$
	Army Army								
	AGCT	Educ. Trng.	Exp.	-1	Total	Matrix A'			
Matrix A									Total
AGCT	$V_1$ 1	.481	.146	.118	-.1	.745	.481	.146	.118
Education	$V_2$ .481	1	.134	-.119	-.1	.496	.231	.070	.057
Army Training	$V_3$ .146	.134	1	-.102	-.1	.178	.070	.021	.017
Army Experience	$V_n$ .118	-.119	-.102	1	-.1	-.103	.057	.017	.014
Test Score	$C$ .490	.319	.315	.329	0	1.453	.236	.072	.058
	$D$ .769	.064	-.176	.481	-.1	.138	Matrix B'		
Matrix B	1	.083	-.229	.626	-.1	.480	.083	-.229	.626
	.064	.979	-.119	.146	-.1	.069	.005	-.015	.040
	-.176	-.119	.986	.118	-.1	-.191	-.015	.040	-.110
	.083	.243	.271	.490	0	1.088	.007	-.019	.052
	$D'$ .974	-.104	.106	.064	-.1	.038	Matrix C'		
Matrix C	1	-.107	.109	.066	-.1	.068	-.107	.109	.066
	-.104	.946	.228	-.176	-.1	-.106	.011	-.011	-.007
	.236	.290	.438	.083	0	1.048	-.025	.026	.016
	$D''$ .935	.239	-.169	-.104	-.1	-.099	Matrix D'		
Matrix D	1	.256	-.181	-.111	-.1	-.036	.256	-.181	-.111
	.315	.412	.067	.236	0	1.032	.081	-.057	-.035
	$D'''$ .331	.124	.271	.315	1.043		Matrix D'		
$\beta$ Values	.331	.161	.278	.337					
$\beta$ times $r$	.162	.051	.088	.111	.412	$R_{c,1234} = .64$			

TABLE 2  
Proof of the  $\beta$  values by Formula  
 $r_{c1} = \beta_1 + \beta_2 r_{12} + \beta_3 r_{13} + \beta_4 r_{14}$ , etc.

$\beta$	$V_1$	$V_2$	$V_3$	$V_n$	$V_1\beta_i$	$V_2\beta_i$	$V_3\beta_i$	$V_4\beta_i$
.331	1	.481	.146	.118	.331	.159	.048	.039
.161	.481	1	.134	-.119	.077	.161	.022	-.019
.278	.146	.134	1	-.102	.041	.037	.278	-.028
.337	.118	-.119	-.102	1	.040	-.040	-.034	.337
$r$ with criterion	.490	.319	.315	.329				
Totals					.489	.317	.314	.329
Differences	.001	.002	.001	.000				

dition throughout the problem, as just described for Matrix  $A'$ . Small discrepancies in the last decimal place may result from dropping additional decimal places, but if gross discrepancies are found, an error has been made in computation and must be corrected before proceeding further.

Another check that should be made for each of the matrices on the right-hand side of the page is to see that the part of each matrix that is supposed to be symmetrical is so. If not, an error has been made which of course must be corrected before continuing with the problem.

*Obtaining Values for Row D.* Row  $D$  is an intermediate row between Matrix  $A$  and Matrix  $B$ , and is obtained by subtracting in row  $V_2$  the numbers in Matrix  $A'$  from the corresponding numbers in Matrix  $A$ . From the 1 in row  $V_2$ , column  $V_2$  is subtracted the .231 in row  $V_2$ , column  $V_2$ . The resulting .769 is placed in column  $V_1$  of the  $D$  row. Continuing the subtractions in row  $V_2$ , from .134 is subtracted .070, and the difference .064 is placed in the  $D$  row; from  $-.119$  is taken .057, and the algebraic difference,  $-.176$ , occupies the next place in the  $D$  row. Since there is nothing to subtract from the  $-1$  in row  $V_2$ , the .481 from the first column of the row is brought down into the next column of the  $D$  row.  $A - 1$  is entered in the  $-1$  column, and in the  $T$  column the .138 is the difference between .496 and .358 in the  $T$  and  $T'$  columns of row  $V_2$ . The  $D$  row should be checked at this point to see that the number in the  $T$  column is the algebraic sum of the row.

*Matrix B.* The numbers obtained in row  $D$  should be divided by the first number in the row (.769 in this case) and checked in the usual manner by means of the  $T$  column. In this manner a 1 always appears in the upper left-hand corner of Matrix  $B$ , which will reduce the arithmetic considerably when solving this matrix. Let the first  $n$  numbers as obtained from this division in the first row stand, but eliminate the last two, and substitute  $-1$  in the  $-1$  column and in column  $T$  enter the algebraic sum of the row as now constituted. This is the first row of Matrix  $B$ .

The second row of Matrix  $B$  is obtained from row  $V_3$  of Matrices  $A$  and  $A'$  by subtracting the numbers in Matrix  $A'$  from the corresponding numbers in Matrix  $A$  (in the sample  $.134 - .070 = .064$ ;  $1 - .021 = .979$ , etc.). Then enter  $-1$  in the  $-1$  column, and in the  $V_n$  column the number from row  $V_3$ , column  $V_1$  of Matrix  $A$  (.146 in the sample). Check as usual by addition.

Build each of the  $n$  rows of Matrix  $B$  in the same manner, using successive rows of Matrices  $A$  and  $A'$ . The  $V_n$  column of Matrix  $B$  (except in the first row) contains the same numbers in the same

order as the  $V_1$  column of Matrix  $A$  (except the first two). The last row of Matrix  $B$  has a 0 in the  $-1$  column the same as Matrix  $A$ .

It will be noted that in Matrix  $B$  there is a symmetrical square of  $n-2$  arrays, starting in the second row and column of the matrix. This reduces the number of subtractions necessary, since some numbers can be copied into their proper places, as soon as one sees where the symmetry lies. (In the example these numbers in Matrix  $B$  are .064,  $-.176$ , and  $-.119$ )

*Matrix  $B'$ .* When the  $n$  rows of Matrix  $B$  are completed and checked, make a table of products in Matrix  $B'$ , exactly as was done for Matrix  $A$ . Check the algebraic sum of each row as before. There will be a symmetrical submatrix of  $n-2$  arrays contained within Matrix  $B'$ . (In the example the only symmetrical number is  $-.015$  since there are only four independent variables in the problem.) If this square is not symmetrical, an error has been made and must be corrected before proceeding further. (The error may be in the first row of Matrix  $B$ , where all the numbers were divided by a constant.)

*Row  $D'$ .* The differences below Matrix  $B$ , in row  $D'$  are obtained from the second row of Matrices  $B$  and  $B'$  in the same way that row  $D$  was obtained from row  $V_2$  of Matrices  $A$  and  $A'$ . (In the sample  $.979 - .005 = .974$ ;  $-.119 - (-.015) = -.104$ , etc.). The usual check by addition should be made.

*Matrix  $C$ .* Build Matrix  $C$  from Matrix  $B$  exactly as Matrix  $B$  was obtained from Matrix  $A$ . There will be  $n-1$  rows in Matrix  $C$ .

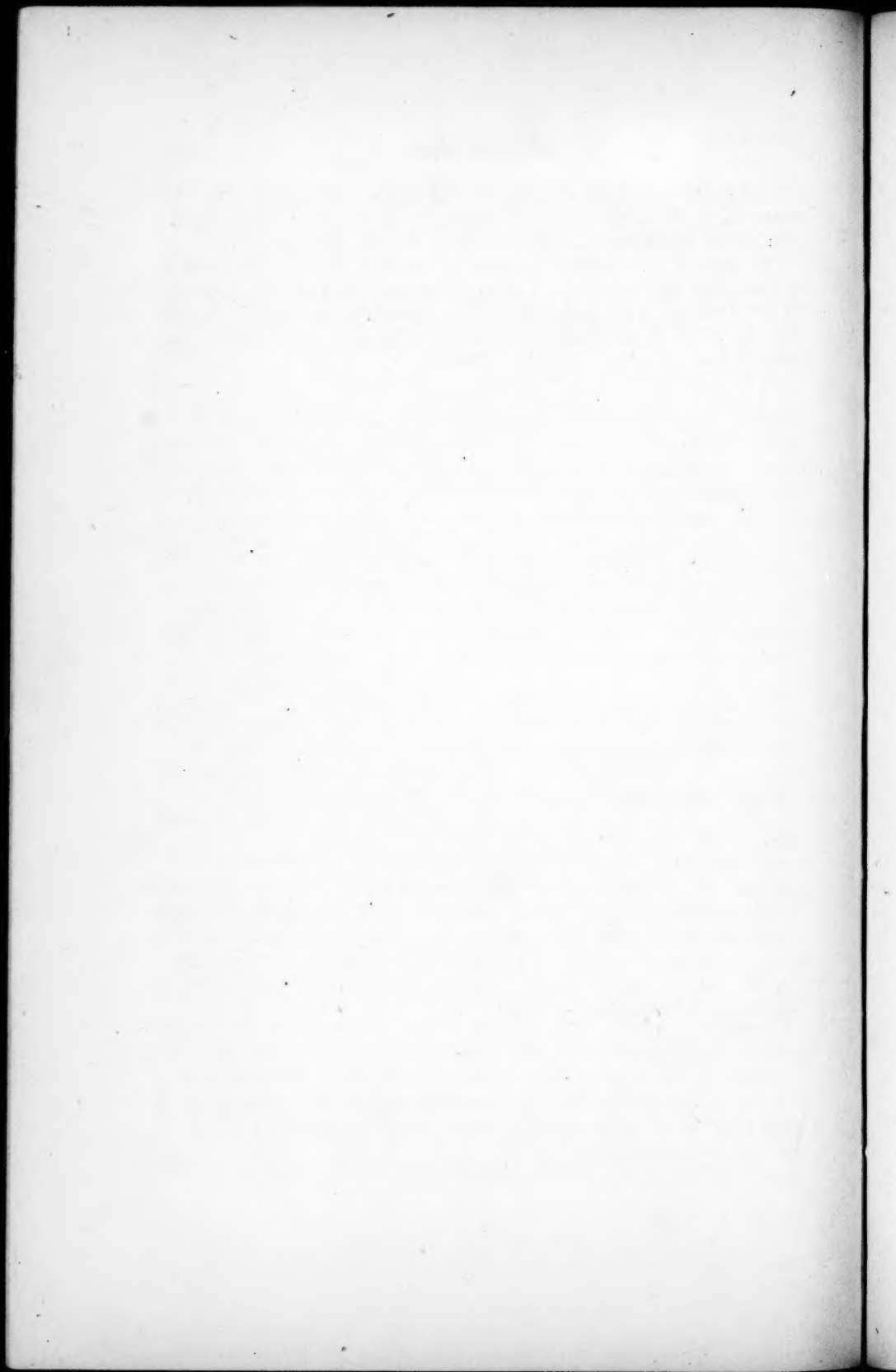
*Matrix  $C'$ .* Matrix  $C'$  is derived from Matrix  $C$  exactly as  $B'$  was derived from  $B$  and  $A'$  from  $A$ .

Continue building matrices until the last contains but 2 rows. (This will be the  $n$ th matrix).

From this last matrix the beta values are obtained. First, a  $D$  row is obtained by subtracting the products obtained at the right from the numbers in the second row of the matrix. In the sample these differences are .331, .124, etc., in row  $D''$ . The number thus obtained in the  $V_1$  column (.331) is the first beta. The second beta is obtained by dividing the number at the foot of the  $V_2$  column (.124) by the number that row 1 of Matrix  $B$  was divided by ( $.124 \div .769 = .161$ ). (This divisor is  $1 - r^2$ ). The third beta is obtained by dividing the number at the foot of the  $V_3$  column (.271) by the number that row 1 of Matrix  $C$  was divided by (.974). The fourth beta is similarly obtained by dividing the number at the foot of the  $V_4$  column by the number that row 1 of Matrix  $D$  was divided by ( $.315 \div .935 = .337$ ). For problems containing more variables, continue dividing the numbers obtained from the last matrix by the numbers that the top rows of each corresponding matrix were divided by.

Multiply each beta value by the respective correlation with the criterion, in the order entered in the  $C$  row of Matrix  $A$ . These products should be added; the square root of the sum is multiple  $R$ .

A sample five-variable problem is solved in Table 1. In Table 2 is presented the proof of the beta values for this example, obtained by substitution in the standard formula quoted in the table. It is seen that the totals agree closely with the coefficients of correlation of the independent variables with the criterion.



DYNAMICS OF THE CEREBRAL CORTEX AUTOMATIC  
DEVELOPMENT OF EQUILIBRIUM IN  
SELF-ORGANIZING SYSTEMS

W. R. ASHBY

DEPARTMENT OF RESEARCH, BARNWOOD HOUSE  
GLOUCESTER, ENGLAND

The nervous system, and particularly the cerebral cortex, is examined in certain of its activities, it being treated as a purely physical dynamic system. It is shown mathematically that if, as seems likely, the cerebral cortex can undergo manifold changes of neuronal organization, then it follows that these changes must automatically lead to the development of more and more states of equilibrium, the process being unavoidable and largely irreversible. The theorem may be established on basic physical principles without appeal to special physiological details. The possible significance of this observation is indicated.

1.

It is generally accepted that the nervous system may be regarded as a purely physico-chemical or mechanistic system. This involves two assumptions:

(a) That its state may be specified by the numerical values at any moment of suitably selected variables. (The technical difficulty of actual measurement and the great number of them required do not affect the principle);

(b) That if we knew the state of every reacting part we could predict with certainty what it would do next; or in other words, its change at any moment depends only on its present state. (It will commonly be necessary to include environmental actions and reactions to make the system complete.)

This means (Note 1)\* that the behavior of the system can be described by equations of the type

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \dots, x_n) \quad (i = 1, 2, \dots, n). \quad (1)$$

Such a system of variables, i.e., whose fluxions are functions only of the variables of that system,  $t$  in particular being absent from the

\* To avoid interrupting the line of argument with minor, but necessary, matters, these have been collected into "Notes" at the end of the paper.

right-hand side, is an "absolute" system. It is a well-known theorem that such a system of equations defines a congruence of curves in the corresponding  $n$ -dimensional space having co-ordinates  $x_1, x_2, \dots, x_n$ . This congruence, the " $x$ -field," forms a representation of the behavior of the system which is very convenient for the later proofs (Note 2).

The "organization" of the system must now be identified with the set of  $f$ 's of equations (1). No other identification is possible.

It appears probable that some parts of the nervous system, particularly the cerebral cortex, can undergo internal reorganizations which result in a change of behavior. The many facts of learning and training indicate this. In view of the known variability of behavior in the higher animals, it seems that this postulate must be allowed; for if we hold to the "mechanistic" hypothesis we must allow that changed behavior must be the result of some inner, physical change (Note 3).

These "reorganizations" are the subject of this paper. It will be shown that they must always tend in the direction of increasing equilibrium.

## 2.

It will now be shown that a *spontaneous change of organization implies the presence of a step-function of the time*. A change of organization, by definition (§ 1), means that the functional forms  $f_i$  must change to  $\phi_i$ , say. This change may be represented with equal generality as being due to a parameter  $h$ , in the  $f$ 's, the  $h$  changing value from, say,  $h'$  to  $h''$ . This gives the two required organizations, i.e.,

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \dots, x_n; h')$$

and

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \dots, x_n; h''). \quad (2)$$

This change (from  $h'$  to  $h''$ ), however, by the "absoluteness" hypothesis (equation 1), is not to be arbitrary but is to depend on the internal events of the system. This means that  $h$  is to be treated as an extra variable in the system and that its change must be a function of given  $x, h$  combinations.  $h$ , therefore, regarded as an ordinary variable of the system, will be observed in its time-development to be constant at  $h'$  for a time, and then to change to  $h''$ , where it continues.  $h$  must therefore be a step-function of the time (Note 4).

At this point it must be noted that the system composed of both

the  $x$ 's and  $h$  is absolute and has one field in  $n+1$  dimensions; and it has one and only one organization. This system must be clearly distinguished from the system composed of the  $x$ 's only, which is also absolute in the intervals while  $h$  is constant, and in this interval it has one field in  $n$  dimensions. But owing to the possibility of  $h$ -change, the  $x$ -system has *two possible organizations*. The effect of the change of  $h$  is to change one  $x$ -field to another (Equations 2) (Note 5). This sudden change of a field from one form to another due to movement of the  $x$ -point in the field is fundamental here, and the concept will be used repeatedly in the paper. It is useless to proceed unless the reader is satisfied with the validity of this concept.

The presence of more than one step-function in the system adds further possible spontaneous changes of organization in the  $x$ -system. If there are  $s$  step-functions  $h$ , each of which may take  $r$  values, then the  $x$ 's will have  $r^s$  organizations. (Each step-function must have defined the  $x$  and  $h$  points at which it changes value. These points will be referred to as  $\theta$ -points, from the equation in Note 4).

We now define a *commutative* system as one which ( $\alpha$ ) is absolute, ( $\beta$ ) contains step-functions of the time  $h_i$ , ( $\gamma$ ) contains an infinite number of organizations of the variables  $x$ , provided by the step-functions  $h$ .

It is now easily shown that the initial hypotheses of § 1 are equivalent to: *The nervous system and its environment form a commutative system* (Note 6).

### 3.

We now proceed to study a peculiar property of commutative systems. A commutative system being absolute, its variables  $x_i$  and  $h_i$  will change with the time, the  $h$ 's change being discontinuous. (Some further properties are mentioned in Note 7). We now ask: Under what conditions can the  $h$ 's become permanently constant, in spite of repeated disturbances affecting the  $x$ 's? (Note 8) Constancy of the  $h$ 's is important since it corresponds to constancy of the  $x$ -field; and it is therefore the necessary and sufficient condition that the  $x$ 's should demonstrate a constant pattern of behavior after disturbance. An important sufficient condition is given by the following theorem: In any commutative system, if some external disturbance brings the  $x$ -point repeatedly to points  $A, B, \dots$  (at random times and in random order) a sufficient condition (Note 9) for subsequent permanent constancy of the  $h$ 's is that an  $x$ -field should occur with both the  $x$ -point at that moment and the points  $A, B, \dots$  all on paths which (a) after some finite time meet no new  $x$ -points and (b) meet no

$\theta$ -points. *Proof:* On the occurrence of such a field, the  $x$ -point may continue to move as long as we please (Note 10) but no  $\theta$ -points can be met by the  $x$ -point, since it can only follow the paths from  $A, B, \dots$  and these contain no  $\theta$ -points. And as change of field occurs only on the  $x$ -point meeting a  $\theta$ -point, change of field and of path cannot occur; so the  $x$ -point is permanently kept to these paths and is therefore permanently barred from meeting a  $\theta$ -point. So the  $h$ 's and the field are permanently constant.

A definition of "equilibrium" is now required. A path is defined here as "equilibrical" if, from some point onwards, it always remains within a fixed region of the  $x$ -space (Note 11). It is clear that the paths mentioned above (which meet no new  $x$ -points after a finite time) are equilibrical in the sense given. For after the finite time, the path must either come to an end, or join back on to itself forming a cycle; but in either case it remains within a finite region.

As there are an infinite number of fields we now proceed by using probability methods (Note 7). It will now be shown that: if there is a constant probability  $p$  that a field provided by a random combination of  $h$ -values in a commutative system will be of the type proved sufficient for  $h$ -constancy above (Note 12), then the probability  $P$  that the system after  $m$  field-changes will be  $h$ -constant is at least (Note 13)

$$1 - (1 - p)^m.$$

*Proof:* The occurrence of such a field being sufficient to cause  $h$ -constancy (here called "success"), and since a success stays so, while the failures at each field-change are given a fresh chance  $p$  of succeeding, we have (if  $1 - p = q$ ) that at the first trial  $p$  succeed and  $q$  fail; in the second trial the  $q$  will be divided so that  $qp$  succeed and  $qq$  fail; and so on. The proportion succeeding at the  $m$ -th trial will be  $pq^{m-1}$ , and  $P$  will be at least

$$p + pq + pq^2 + \dots + pq^{m-1} = 1 - (1 - p)^m. \quad (3)$$

In other words,  $m$  field-changes in a commutative system increase the probability of equilibrium from  $p$  to  $1 - (1 - p)^m$ , the latter always being greater. If we now add the trifling postulate that so long as the  $h$ 's are not constant  $m \rightarrow \infty$  as  $t \rightarrow \infty$ , then clearly

$$\lim_{t \rightarrow \infty} P = 1. \quad (4)$$

4.

Equation (4) is the point of the whole paper. It means that, in a commutative system, however small  $p$  may be, making  $m$  (or  $t$ ) large

enough will ensure that  $P$  approaches 1. In other words, no matter how rare equilibrium paths are in the individual  $x$ -fields, a commutative system will inevitably keep changing them until it does develop an equilibrium (Note 14). A commutative system is selective for organizations having equilibrium. And by our previous conclusion (that the nervous system and its environment form a commutative system) this must apply equally to the nervous system.

We see therefore that the assumption that the nervous system is a physical system and that it can undergo changes of organization leads without any further hypothesis (Note 15) to the deduction that it must be a commutative system (§ 2), and this leads to the conclusion (§ 3) that it must tend to develop an organization giving equilibrium, no other organization being capable of persisting.

## 5.

This paper is hardly concerned with applications, but two may be mentioned.

After the establishment of equilibrium has occurred, all disturbances (i.e., displacements towards  $A, B, \dots$ ) will be followed by activities of the system which will always oppose the disturbance (this being a characteristic of all systems in equilibrium).

This opposition to disturbance may, if the activity of the reaction is small, explain the phenomenon of "habituation."

If the reaction to the disturbance is more active and complex, it may explain the development of "adaptive" behavior, since adaptation seems to be essentially the development of that organization which will preserve equilibrium (Notes 16, 17, 18, 19).

## NOTES

- (1) This can be proved in much finer detail, but the details are hardly required here. Its derivation from (b) is almost direct.
- (2) A configuration of the system corresponds to a point in the space. The behavior of the system in time corresponds to the point moving along a path. All the paths make the field and this defines the organization of the system. "Experimental control" is equivalent to the ability to start the point where we like in the field (by using the  $n$  arbitrary constants in the solution of equation 1).
- (3) We do not need here to know anything of the physiological mechanism underlying such change.
- (4) To preserve the form of equations (1) we may use any continuous approximation to step-function form such as

$$\frac{dh}{dt} = q \left[ \frac{h' + h''}{2} + \frac{h' - h''}{2} \tanh\{q \theta(x_1, \dots, x_n; h)\} - h \right];$$

where  $q$  is positive and large. The function  $\theta$  determines those  $x$  and  $h$  combinations where  $h'$  changes to  $h''$ ; for where  $\theta(x_1, \dots, x_n; h)$  is positive,  $h$  becomes  $h'$ , while where it is negative  $h$  changes to  $h''$ .

- (5) The whole  $x$ -field remains constant until  $h$  changes; then the  $x$ -field changes, in general to something quite different. Consequently, if the  $x$ -point, started from  $x_1^0, x_2^0, \dots, x_n^0$  always followed a particular path, then after change of  $h$  it can follow a *different path* even though started from the same ( $x$ ) point.
- (6) (a) The hypothesis (b) demands absoluteness. (b) The "reorganization" hypothesis demands step-functions, as we have shown above. (c) As we are not concerned with the possibility that the nervous system is limited in its variability of behavior, we may assume this to be infinite in the absence of any reason to the contrary.
- (7) Since a commutative system has an infinite number of  $x$ -fields, it clearly cannot be handled with the explicitness of equations (1). Assuming we have in front of us a real, physical example of a commutative system, we can clearly no longer know (a) the number of  $h$ 's, (b) their values, (c) the forms of the  $f$ 's in  $f_i(x_1, \dots; h_1, \dots)$ , (d) the  $x$ -fields, since these depend on the  $f$ 's. But, and this is sufficient for the rest of the paper, (a) we can observe the values and changes of the  $x$ 's, (b) we can control the  $x$ 's starting point, and (c) we can test for constancy of the  $h$ 's by seeing whether the  $x$ 's path in the  $x$ -space repeats itself.
- (8) The disturbances are needed to test for, and demonstrate, equilibrium.
- (9) It is not necessary, for there are other ways of getting  $h$ -constancy, depending on some relation between the distribution of  $\theta$ -points and the direction of paths in the fields. We have no right to say that these combinations cannot occur.
- (10) Some minor postulate that the disturbance is to return the  $x$ -point to  $A, B, \dots$  by some definite route devoid of  $\theta$ -points is required here.
- (11) Firstly, by selecting special examples it is easy to show that "equilibrium" belongs strictly to a single path and not to a field. Other special cases show that the path is to be confined to a region and not necessarily to terminate at a point. It may easily be shown that the common examples of equilibrium are all special cases, or limits, of the definition given. Nothing less general seems to be adequate.
- (12) The  $x$ -point is first to be fixed, but  $p$  is not to depend on where it is fixed.
- (13) "At least" because, as shown in Note 9, other ways of getting  $h$ -constancy are possible, though rare and of little importance.  $p$  does not include these, so their presence may result in an increase of the proportion which have become  $h$ -constant.
- (14) This does not apply to a single field, i.e., the "commutative" part is necessary; for in a single field, as  $t \rightarrow \infty$ ,  $P$  does not  $\rightarrow 1$  but  $\rightarrow p$ . If  $p$  is small the effect is quite different.
- (15) The hypotheses about  $p$  and  $m$  are really only that (1)  $p \neq 0$  and (2)  $m \rightarrow \infty$  as  $t \rightarrow \infty$ . These merely exclude peculiar cases.
- (16) If the  $\theta$ -points are distributed systematically, the paths of equilibrium will automatically have systematic properties imposed on them. This leads to interesting and important developments which have already been explored by the author, but they are outside the scope of this paper.
- (17) The above discussion deals with one system where the  $x$ -point either is, or is not, on an equilibrium path, and the equilibrium is therefore "all or none." But there are several ways in which we can get independent equilibria, and then, following the same principles as before, the number of equilibria must tend to increase by accumulation. But this leads beyond the scope of the present paper.
- (18) We have supposed the environment to remain constant as far as its organization is concerned (though not the values of its variables). This means that the environment can interact freely with the nervous system in the above theory. A single change of the environmental organization would, however, wreck an established equilibrium. But regular changes between a finite set of environments can be shown to tend to equilibrium. If the environment should change its organization irregularly, the whole paper becomes inapplicable since postulate (b) of § 1 is no longer true.
- (19) Ashby, W. R. *J. ment. Sci.*, 1940, 86, 478.

## TABLE OF MEAN DEVIATES FOR VARIOUS PORTIONS OF THE UNIT NORMAL DISTRIBUTION

HOLLIS M. LEVERETT\*  
HARVARD UNIVERSITY

In the absence of sound units of measurement, the data obtained by use of a psychometric instrument may sometimes be so irregular that they cannot be accepted as a good representation of the true function of the trait measured. Frequently, ordered categorical or ranked data and other irregular distributions are obtained to represent traits that are obviously continuous and probably unskewed functions. In such cases, the data may be improved in the direction of a better approximation of the true function by converting them to normal distribution form. A table of mean deviates of various portions of the unit normal curve is presented to facilitate this conversion.

### 1. *Normalized Data*

The practice of normalizing a body of data is not unusual in psychometrics. In the absence of sound measuring instruments, the ordered categorical or ranked data obtained to represent certain functions are frequently converted to a continuous scale and distribution form by the use of the normal curve. A less frequent but equally justifiable practice involves the conversion of continuous but distorted data so that certain corrections are applied to both the scale and the form of the distribution.

Unfortunately, there is a tendency to believe that, when normalizing a body of data, one must be prepared to assert and justify the assumption that the true function represented by the data is a normal curve. Actually, the assumption of a normal curve function is unnecessary. In the absence of sound units of measurement and in the absence of knowledge of the exact mathematical function that represents the true distribution of a mental trait, any obtained psychometric data must (1) be accepted as a sound representation of the scale and distribution of the trait, or (2) be adjusted to fit a more probable scale and distribution. If a measuring instrument is sound and the distribution obtained is reasonable, the data may be accepted as a satisfactory representation of the mental trait. On the other hand, if a measuring instrument is not entirely sound and the distribution obtained is not reasonable, some corrections may be applied to improve the scale and distribution form. In the case that the cor-

\* The author wishes to acknowledge the kind counsel of Professor Truman L. Kelley in the preparation of this article.

rection applied involves normalizing the data, it is not necessary to hold that the true distribution is the normal curve; the action may be taken soundly on the consideration that the normalized data provide a better approximation of the true function than do the original data.

It is possible that curves other than the normal curve might serve the corrective purpose in providing a better approximation of the true situation; the normal distribution, however, is quite reasonable in most situations, and it has the advantage of being a very convenient practical curve. Its use to adjust certain data is justified if (1) the obtained data cannot be accepted as representing either the scale or the distribution form of the true function and (2) the scale and distribution form of the normal curve are considered a better approximation of the true function than the obtained data. When these two conditions exist, the data may be normalized, and, unless a curve more appropriate than the normal curve is known, the corrections made can be considered to yield the best approximation of the true data.

Fundamentally, therefore, a body of data is normalized to obtain better estimates of true values for the obtained measurements. This converted distribution, however, has other practical advantages. The normalized data possess a sound equal-unit linear scale, and the original defective distribution is modified in the direction of all the essential characteristics of the normal curve, a distribution wherein all the algebraic relationships are simple, very tractable, and well known.

## 2. *Tables of Deviates*

A number of tables are available to assist in the conversion of data to normal distribution form. Nearly all of these tables, however, require the use of median deviates for various portions of the curve. When a large number of curves are treated, the computation of accumulative proportions for the median of each class of the data to which a deviate is to be assigned, is awkward and laborious. Furthermore, the mean deviate may be considered the more appropriate statistic for the purpose.

A table of mean deviates published by Thorndike (3) is not satisfactory in several respects. It provides deviate values for only one half of the curve, and all values are given to the second decimal only. In order to obtain the mean deviate for any portion extending over the .50 division of the curve, it is necessary to look up two values and compute a weighted average. This procedure is not only inefficient, but the computation of a weighted mean is hazardous in view of

the two-decimal accuracy of the table values. For the same reason, it is inadvisable to make any interpolations.

The table of mean deviates for various portions of the unit normal curve presented here provides values accurate to the fourth decimal, and the deviates for each .01 portion or multiple thereof can be read directly from the table. In most cases, the hundredths proportions of the data to be normalized will yield sufficient accuracy in the deviates, but cautious linear interpolation is possible if more precise values are needed.

The mean deviate values given in Thorndike's table agree generally with those of the new table. It should be noted, however, that Thorndike's deviates for the first .01 to .10 portions of the curve are too large.

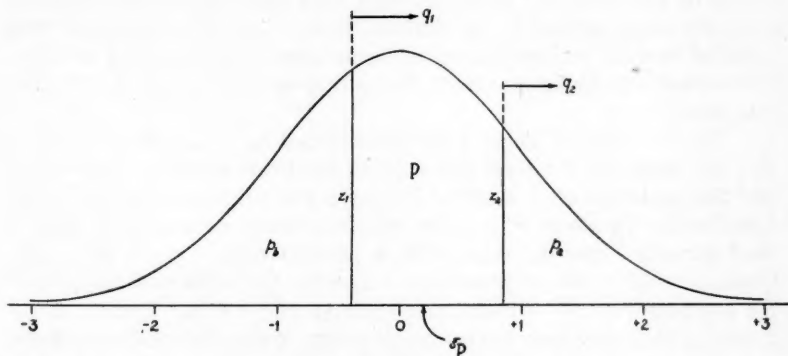


FIGURE 1

### 3. Table of Mean Deviates

Figure 1 shows the variables used to compute the mean deviations for various portions of the unit normal curve and the arguments needed to obtain mean deviates from the table. In this figure,  $P$  represents any portion of the unit normal curve for which the mean deviate  $\bar{s}_p$  is desired.  $z_1$  and  $z_2$  are the ordinates limiting  $P$ ;  $q_1$  and  $q_2$  are the proportions of the curve to the right of  $z_1$  and  $z_2$ , respectively. The small letter  $p$  is used to designate the portion of the curve that is either above or below  $P$ . Subscripts  $a$  and  $b$  differentiate between  $p_a$ , the portion of the curve above  $P$  including the positive tail, and  $p_b$ , the portion of the curve below  $P$  including the negative tail. The relationships among these variables to be noted are (1) that  $q_1 - q_2 = P$  and (2) that  $p_a + P + p_b = 1$ .

The mean deviates given in Table 1 were computed in accordance with the formula given by Kelley (1). This formula may be written

$$\bar{s}_P = \frac{z_1 - z_2}{q_1 - q_2}.$$

In special cases where  $P$ , the portion of the curve for which the mean deviate is desired, extends to and includes a tail of the distribution, this equation becomes

$$\bar{s}_P = \frac{z}{P}.$$

The ordinate values,  $z_1$  and  $z_2$ , used in computing mean deviates were taken from *The Kelley Statistical Tables* (2). To assure accuracy of the table, the mean deviates were computed twice. Since all numbers were carried to six decimal places and the deviations were rounded to four decimal places in accordance with standard practice, it is reasonable to assume that the table is correct to the fourth decimal place.

The columns of Table 1 are determined by  $p$  values from .00 to .50; the rows, by  $P$  values from .01 to the limit where  $p + \frac{1}{2}P = .50$  and the deviation of  $P$  is zero. Taking  $p$  from the positive end of the distribution ( $p$  above  $P$ ,  $p_a$ ), the required mean deviate of  $P$  may be read directly from the table with a positive sign if  $p_a + \frac{1}{2}P < .50$ . When  $p_a + \frac{1}{2}P > .50$ , it is necessary to enter the table with the  $p$  value for the negative end of the distribution ( $p$  below  $P$ ,  $p_b$ ) and the deviation of  $P$  thus obtained has a negative sign. As a universal rule, therefore, the sign of a deviate read from Table 1 is that of the tail of the distribution included in  $p$ .

#### 4. Use of Table 1

Table 1 facilitates the conversion of any distribution to normal curve form. For demonstration purposes, we shall normalize a distribution in which the frequencies of successive categories are 25, 50, 150, 75, and 200. The conversion is shown in Table 2, and the steps in the process are as follows:

1. The categories are entered in the worksheet in positive to negative order as shown in column A.
2. The frequency in each category and the total,  $N$ , are entered in column B.
3. The frequency in each category is divided by  $N$  to obtain the  $P$  values in column C.

4. The proportion of the distribution above each category,  $p_a$ , is obtained by summing  $P$  values. In the example shown,  $p_a$  is obtained for categories 1 through 3. Since for category 4,  $p_a + \frac{1}{2}P > .50$ , the procedure is reversed and  $p_b$ , the proportion of the distribution below each category, is obtained for categories 4 and 5. These  $p$  values are entered in column D.
5. The mean deviation of each category is obtained by entering Table 1 with the  $p$  and  $P$  value for that category. These deviates are shown in column E. When  $p_a$  is used (categories 1-3) the deviates are positive; when  $p_b$  is used (categories 4-5) the deviates are negative.

TABLE 1

Mean Deviates of Various Portions of the Unit Normal Curve

 $p$  = the proportion of the curve above or below $P$  = the portion for which the mean deviate is required $p = .00-.09$   
 $P = .01-.50$ 

$p$ of distribution above or below $P$										
$P$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.01	2.6652	2.1766	1.9624	1.8132	1.6962	1.5987	1.5145	1.4398	1.3725	1.3107
.02	2.4209	2.0695	1.8878	1.7549	1.6474	1.5566	1.4772	1.4062	1.3416	1.2822
.03	2.2681	1.9841	1.8239	1.7027	1.6031	1.5177	1.4423	1.3743	1.3123	1.2550
.04	2.1544	1.9121	1.7676	1.6556	1.5623	1.4814	1.4094	1.3442	1.2844	1.2289
.05	2.0627	1.8494	1.7170	1.6125	1.5243	1.4472	1.3783	1.3154	1.2576	1.2037
.06	1.9854	1.7936	1.6708	1.5725	1.4887	1.4150	1.3486	1.2880	1.2318	1.1795
.07	1.9181	1.7431	1.6282	1.5351	1.4552	1.3843	1.3203	1.2616	1.2070	1.1560
.08	1.8583	1.6967	1.5885	1.4999	1.4233	1.3551	1.2932	1.2361	1.1831	1.1333
.09	1.8043	1.6538	1.5513	1.4666	1.3930	1.3271	1.2671	1.2116	1.1599	1.1112
.10	1.7550	1.6138	1.5162	1.4350	1.3640	1.3002	1.2419	1.1879	1.1373	1.0897
.11	1.7094	1.5763	1.4830	1.4049	1.3362	1.2743	1.2176	1.1648	1.1154	1.0688
.12	1.6670	1.5408	1.4513	1.3760	1.3095	1.2493	1.1940	1.1425	1.0941	1.0484
.13	1.6273	1.5071	1.4211	1.3482	1.2837	1.2251	1.1711	1.1207	1.0733	1.0285
.14	1.5898	1.4750	1.3921	1.3215	1.2588	1.2016	1.1488	1.0995	1.0531	1.0090
.15	1.5544	1.4444	1.3642	1.2957	1.2346	1.1788	1.1272	1.0788	1.0332	.9899
.16	1.5207	1.4150	1.3374	1.2708	1.2112	1.1567	1.1061	1.0586	1.0138	.9712
.17	1.4886	1.3868	1.3115	1.2466	1.1884	1.1350	1.0854	1.0389	.9948	.9528
.18	1.4578	1.3595	1.2864	1.2231	1.1662	1.1140	1.0653	1.0195	.9761	.9348
.19	1.4282	1.3332	1.2620	1.2003	1.1446	1.0934	1.0455	1.0005	.9578	.9170
.20	1.3998	1.3077	1.2384	1.1780	1.1235	1.0732	1.0262	.9819	.9398	.8996
.21	1.3724	1.2831	1.2154	1.1563	1.1029	1.0535	1.0073	.9636	.9221	.8824
.22	1.3459	1.2591	1.1930	1.1352	1.0827	1.0341	.9886	.9456	.9047	.8655
.23	1.3202	1.2357	1.1711	1.1144	1.0629	1.0152	.9704	.9279	.8875	.8488
.24	1.2953	1.2130	1.1498	1.0942	1.0435	.9965	.9524	.9105	.8706	.8323
.25	1.2711	1.1909	1.1289	1.0743	1.0245	.9782	.9347	.8934	.8539	.8161
.26	1.2476	1.1692	1.1085	1.0549	1.0058	.9602	.9173	.8765	.8375	.8000
.27	1.2246	1.1480	1.0885	1.0357	.9875	.9425	.9001	.8598	.8212	.7842
.28	1.2022	1.1273	1.0688	1.0170	.9694	.9250	.8832	.8433	.8052	.7685
.29	1.1804	1.1070	1.0496	.9985	.9516	.9078	.8655	.8271	.7893	.7530
.30	1.1590	1.0871	1.0306	.9804	.9341	.8909	.8500	.8110	.7736	.7376
.31	1.1380	1.0676	1.0120	.9625	.9169	.8741	.8337	.7951	.7581	.7224
.32	1.1175	1.0484	.9937	.9449	.8998	.8576	.8176	.7794	.7427	.7074
.33	1.0974	1.0296	.9757	.9275	.8830	.8413	.8017	.7639	.7275	.6924
.34	1.0777	1.0110	.9579	.9104	.8664	.8251	.7859	.7485	.7124	.6776
.35	1.0583	.9928	.9404	.8935	.8500	.8092	.7704	.7332	.6975	.6629
.36	1.0392	.9748	.9232	.8768	.8338	.7934	.7549	.7181	.6827	.6484
.37	1.0205	.9570	.9061	.8603	.8178	.7777	.7396	.7031	.6679	.6339
.38	1.0020	.9395	.8893	.8440	.8019	.7622	.7245	.6883	.6533	.6195
.39	.9838	.9223	.8726	.8278	.7862	.7469	.7094	.6735	.6388	.6053
.40	.9659	.9052	.8562	.8119	.7706	.7317	.6945	.6589	.6244	.5911
.41	.9482	.8884	.8399	.7960	.7552	.7166	.6797	.6443	.6101	.5770
.42	.9307	.8717	.8238	.7804	.7399	.7016	.6650	.6299	.5959	.5629
.43	.9135	.8553	.8079	.7649	.7247	.6868	.6505	.6155	.5817	.5490
.44	.8964	.8390	.7921	.7495	.7097	.6720	.6360	.6012	.5677	.5351
.45	.8796	.8229	.7764	.7342	.6948	.6573	.6215	.5871	.5537	.5212
.46	.8629	.8069	.7609	.7191	.6799	.6428	.6072	.5729	.5397	.5074
.47	.8464	.7910	.7455	.7040	.6652	.6283	.5930	.5589	.5258	.4937
.48	.8301	.7753	.7303	.6891	.6506	.6139	.5788	.5449	.5120	.4800
.49	.8139	.7598	.7151	.6743	.6360	.5996	.5647	.5309	.4982	.4663
.50	.7979	.7443	.7000	.6595	.6215	.5853	.5506	.5170	.4845	.4527



TABLE 1 (continued)

 $p = .10-.19$   
 $P = .01-.50$ 

<i>p</i> of distribution above or below <i>P</i>										
<i>P</i>	.10	.11	.12	.13	.14	.15	.16	.17	.18	.19
.01	1.2538	1.2004	1.1505	1.1032	1.0582	1.0153	.9742	.9346	.8965	.8597
.02	1.2271	1.1754	1.1268	1.0807	1.0368	.9948	.9544	.9156	.8781	.8418
.03	1.2016	1.1514	1.1040	1.0589	1.0159	.9747	.9351	.8969	.8600	.8243
.04	1.1770	1.1281	1.0818	1.0377	.9956	.9552	.9162	.8787	.8423	.8071
.05	1.1532	1.1055	1.0603	1.0171	.9758	.9361	.8978	.8608	.8250	.7902
.06	1.1302	1.0836	1.0393	.9970	.9564	.9174	.8797	.8432	.8079	.7735
.07	1.1079	1.0623	1.0189	.9774	.9375	.8991	.8619	.8260	.7911	.7571
.08	1.0863	1.0416	.9990	.9582	.9190	.8811	.8445	.8090	.7746	.7410
.09	1.0652	1.0214	.9796	.9394	.9008	.8635	.8274	.7923	.7583	.7251
.10	1.0446	1.0017	.9605	.9210	.8830	.8462	.8105	.7759	.7422	.7094
.11	1.0246	.9823	.9419	.9030	.8655	.8291	.7939	.7597	.7264	.6939
.12	1.0050	.9634	.9236	.8853	.8482	.8124	.7776	.7438	.7108	.6786
.13	.9858	.9449	.9057	.8678	.8313	.7959	.7615	.7280	.6953	.6634
.14	.9670	.9267	.8880	.8507	.8146	.7796	.7456	.7124	.6801	.6485
.15	.9485	.9089	.8707	.8339	.7982	.7636	.7299	.6970	.6650	.6336
.16	.9304	.8913	.8536	.8172	.7820	.7477	.7144	.6818	.6501	.6190
.17	.9126	.8740	.8368	.8009	.7660	.7321	.6990	.6668	.6353	.6044
.18	.8951	.8570	.8203	.7847	.7502	.7166	.6839	.6519	.6207	.5900
.19	.8779	.8403	.8040	.7688	.7346	.7013	.6689	.6372	.6062	.5757
.20	.8610	.8238	.7879	.7530	.7192	.6862	.6540	.6226	.5918	.5616
.21	.8443	.8075	.7719	.7375	.7039	.6712	.6393	.6081	.5775	.5475
.22	.8278	.7914	.7562	.7221	.6888	.6564	.6247	.5938	.5634	.5336
.23	.8115	.7755	.7407	.7068	.6739	.6417	.6103	.5795	.5493	.5197
.24	.7955	.7598	.7253	.6918	.6591	.6272	.5960	.5654	.5354	.5059
.25	.7796	.7443	.7101	.6768	.6444	.6127	.5817	.5514	.5215	.4922
.26	.7639	.7290	.6951	.6621	.6299	.5984	.5676	.5374	.5078	.4786
.27	.7484	.7138	.6801	.6474	.6154	.5842	.5536	.5236	.4941	.4651
.28	.7331	.6987	.6654	.6329	.6011	.5701	.5397	.5098	.4805	.4516
.29	.7179	.6838	.6507	.6184	.5869	.5561	.5258	.4961	.4669	.4382
.30	.7028	.6690	.6362	.6041	.5728	.5422	.5121	.4825	.4535	.4248
.31	.6879	.6544	.6218	.5899	.5588	.5283	.4984	.4690	.4401	.4115
.32	.6731	.6398	.6074	.5758	.5449	.5145	.4848	.4555	.4267	.3983
.33	.6584	.6254	.5932	.5618	.5310	.5009	.4712	.4421	.4134	.3851
.34	.6439	.6111	.5791	.5478	.5172	.4872	.4577	.4287	.4001	.3719
.35	.6294	.5969	.5651	.5340	.5035	.4737	.4443	.4154	.3869	.3588
.36	.6151	.5827	.5511	.5202	.4899	.4602	.4309	.4021	.3737	.3457
.37	.6009	.5687	.5372	.5065	.4763	.4467	.4176	.3889	.3606	.3326
.38	.5867	.5547	.5234	.4928	.4628	.4333	.4043	.3757	.3474	.3195
.39	.5726	.5408	.5097	.4792	.4493	.4199	.3910	.3625	.3343	.3065
.40	.5586	.5270	.4960	.4657	.4359	.4066	.3778	.3493	.3212	.2935
.41	.5447	.5132	.4824	.4522	.4225	.3933	.3646	.3362	.3082	.2804
.42	.5308	.4995	.4688	.4387	.4092	.3801	.3514	.3231	.2951	.2674
.43	.5170	.4858	.4553	.4253	.3958	.3668	.3382	.3100	.2820	.2544
.44	.5033	.4722	.4418	.4119	.3825	.3536	.3251	.2969	.2690	.2414
.45	.4896	.4586	.4283	.3986	.3693	.3404	.3119	.2838	.2559	.2283
.46	.4759	.4451	.4149	.3852	.3560	.3272	.2988	.2707	.2429	.2153
.47	.4623	.4316	.4015	.3719	.3428	.3140	.2857	.2576	.2298	.2022
.48	.4487	.4181	.3881	.3586	.3295	.3009	.2725	.2445	.2167	.1891
.49	.4352	.4047	.3748	.3453	.3163	.2877	.2594	.2313	.2036	.1760
.50	.4217	.3913	.3614	.3320	.3031	.2745	.2462	.2182	.1904	.1629

$p = .10-.19$   
 $P = .51-.79$

[illegible]

TABLE 1 (continued)

 $p = .20-.29$   
 $P = .01-.50$ 

<i>p</i> of distribution above or below <i>P</i>										
<i>P</i>	.20	.21	.22	.23	.24	.25	.26	.27	.28	.29
.01	.8239	.7893	.7554	.7225	.6903	.6589	.6281	.5977	.5681	.5389
.02	.8066	.7724	.7390	.7064	.6746	.6434	.6129	.5829	.5535	.5245
.03	.7895	.7557	.7228	.6906	.6591	.6282	.5980	.5682	.5390	.5102
.04	.7728	.7394	.7068	.6749	.6438	.6132	.5832	.5537	.5247	.4961
.05	.7563	.7233	.6910	.6595	.6286	.5983	.5686	.5393	.5105	.4821
.06	.7400	.7074	.6755	.6443	.6137	.5836	.5541	.5250	.4964	.4682
.07	.7241	.6917	.6601	.6292	.5989	.5691	.5398	.5109	.4825	.4545
.08	.7083	.6763	.6450	.6143	.5842	.5547	.5256	.4969	.4687	.4408
.09	.6927	.6610	.6300	.5996	.5697	.5404	.5115	.4830	.4550	.4272
.10	.6773	.6459	.6152	.5850	.5554	.5262	.4975	.4692	.4413	.4138
.11	.6621	.6310	.6005	.5706	.5411	.5122	.4837	.4555	.4278	.4003
.12	.6471	.6162	.5860	.5563	.5270	.4983	.4699	.4419	.4143	.3870
.13	.6322	.6016	.5716	.5421	.5130	.4845	.4563	.4284	.4010	.3738
.14	.6175	.5871	.5573	.5280	.4992	.4707	.4427	.4150	.3876	.3606
.15	.6029	.5728	.5432	.5141	.4854	.4571	.4292	.4017	.3744	.3475
.16	.5885	.5586	.5291	.5002	.4717	.4436	.4158	.3884	.3612	.3344
.17	.5742	.5444	.5152	.4864	.4581	.4301	.4025	.3752	.3481	.3214
.18	.5600	.5304	.5014	.4728	.4446	.4167	.3892	.3620	.3351	.3084
.19	.5459	.5165	.4876	.4592	.4311	.4034	.3760	.3489	.3221	.2955
.20	.5319	.5027	.4740	.4457	.4177	.3901	.3628	.3358	.3091	.2826
.21	.5180	.4890	.4604	.4323	.4044	.3769	.3498	.3228	.2962	.2697
.22	.5042	.4754	.4469	.4189	.3912	.3638	.3367	.3099	.2833	.2569
.23	.4905	.4618	.4335	.4056	.3780	.3507	.3237	.2969	.2704	.2441
.24	.4769	.4483	.4202	.3924	.3649	.3377	.3107	.2840	.2576	.2313
.25	.4634	.4349	.4069	.3792	.3518	.3247	.2978	.2712	.2448	.2185
.26	.4499	.4216	.3936	.3660	.3387	.3117	.2849	.2583	.2320	.2058
.27	.4365	.4083	.3805	.3529	.3257	.2988	.2720	.2455	.2192	.1930
.28	.4231	.3951	.3673	.3399	.3127	.2858	.2592	.2327	.2064	.1803
.29	.4098	.3819	.3542	.3269	.2998	.2730	.2463	.2199	.1937	.1676
.30	.3966	.3687	.3412	.3139	.2869	.2601	.2335	.2071	.1809	.1548
.31	.3834	.3556	.3281	.3009	.2740	.2472	.2207	.1943	.1681	.1421
.32	.3702	.3425	.3151	.2880	.2611	.2344	.2079	.1816	.1554	.1293
.33	.3571	.3295	.3022	.2751	.2482	.2216	.1951	.1688	.1426	.1165
.34	.3440	.3165	.2892	.2622	.2354	.2087	.1823	.1560	.1298	.1037
.35	.3310	.3035	.2763	.2493	.2225	.1959	.1695	.1432	.1170	.0909
.36	.3179	.2905	.2633	.2364	.2096	.1831	.1566	.1303	.1042	.0780
.37	.3049	.2775	.2504	.2235	.1968	.1702	.1438	.1175	.0913	.0652
.38	.2919	.2646	.2375	.2106	.1839	.1573	.1309	.1046	.0784	.0522
.39	.2789	.2516	.2246	.1977	.1710	.1445	.1180	.0917	.0655	.0392
.40	.2660	.2387	.2117	.1848	.1581	.1316	.1051	.0788	.0525	.0262
.41	.2530	.2257	.1987	.1719	.1452	.1186	.0922	.0658	.0394	.0131
.42	.2400	.2128	.1858	.1589	.1322	.1056	.0792	.0527	.0264	
.43	.2270	.1998	.1728	.1460	.1192	.0926	.0661	.0396	.0132	
.44	.2140	.1868	.1598	.1330	.1062	.0796	.0530	.0265		
.45	.2010	.1738	.1468	.1199	.0932	.0665	.0399	.0133		
.46	.1879	.1608	.1337	.1068	.0800	.0533	.0266			
.47	.1749	.1477	.1206	.0937	.0669	.0401	.0134			
.48	.1618	.1346	.1075	.0805	.0536	.0268				
.49	.1486	.1214	.0943	.0673	.0404	.0134				
.50	.1355	.1082	.0811	.0540	.0270					

$p = .20-.29$   
 $P = .51-.59$

[illegible]

$p = .30-.39$   
 $P = .01-.39$

[illegible]

TABLE 1 (continued)

 $p = .40-.49$   
 $P = .01-.19$ 

$p$ of distribution above or below $P$										
$P$	.40	.41	.42	.43	.44	.45	.46	.47	.48	.49
.01	.2404	.2147	.1891	.1637	.1383	.1130	.0879	.0627	.0376	.0125
.02	.2276	.2019	.1764	.1510	.1256	.1004	.0753	.0502	.0250	
.03	.2148	.1892	.1637	.1383	.1131	.0879	.0627	.0376	.0125	
.04	.2020	.1764	.1510	.1257	.1005	.0753	.0502	.0251		
.05	.1892	.1638	.1384	.1131	.0879	.0627	.0376	.0125		
.06	.1765	.1511	.1258	.1005	.0753	.0502	.0251			
.07	.1639	.1385	.1132	.0880	.0628	.0377	.0126			
.08	.1512	.1259	.1006	.0754	.0502	.0251				
.09	.1386	.1133	.0880	.0628	.0377	.0126				
.10	.1260	.1007	.0755	.0503	.0251					
.11	.1134	.0881	.0629	.0377	.0126					
.12	.1008	.0756	.0503	.0252						
.13	.0882	.0630	.0378	.0126						
.14	.0757	.0504	.0252							
.15	.0631	.0378	.0126							
.16	.0505	.0252								
.17	.0379	.0126								
.18	.0253									
.19	.0127									

TABLE 2

<i>A</i> Category	<i>B</i> <i>N</i>	<i>C</i> <i>P</i>	<i>D</i> <i>p</i>	<i>E</i> <i>s<sub>p</sub></i>
			<i>p<sub>a</sub></i>	
1.	25	.05	.00	2.0627
2.	50	.10	.05	1.3002
3.	150	.30	.15	.5422
			<i>p<sub>b</sub></i>	
4.	75	.15	.40	-.0631
5.	200	.40	.00	-.9659
	500	1.00		

## REFERENCES

1. Kelley, Truman L. Statistical method. New York: Macmillan, 1923.
2. Kelley, Truman L. The Kelley statistical tables. New York: Macmillan, 1938.
3. Thorndike, Edward L. Introduction to the theory of mental and social measurements. New York: Teachers College, Columbia University, 1913.

# NOTE ON THE SAMPLING ERROR OF THE DIFFERENCE BETWEEN CORRELATED PROPORTIONS OR PERCENTAGES

QUINN MCNEMAR  
STANFORD UNIVERSITY

Two formulas are presented for judging the significance of the difference between correlated proportions. The chi square equivalent of one of the developed formulas is pointed out.

It is well known that the sampling variance of the difference between two proportions,  $p_1$  and  $p_2$ , based on independent samples is given by

$$\sigma_d^2 = \sigma_{p_1}^2 + \sigma_{p_2}^2 = \frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2}, \quad (1)$$

or by

$$\sigma_d^2 = \frac{pq}{N_1} + \frac{pq}{N_2}, \quad (2)$$

in which  $p_1$  and  $p_2$  as separate estimates of the unknown population parameters are replaced by

$$p = \frac{p_1 N_1 + p_2 N_2}{N_1 + N_2}$$

as a better estimate which is consistent with the to be tested null hypothesis, and therefore to be preferred. It is also known that when  $\sigma_d^2$  is calculated by (2) the value of  $d^2/\sigma_d^2$  equals  $\chi^2$  with one degree of freedom.

There are many situations in which the sampling variance of the difference between two proportions (or percentages) must take into account the fact that the two proportions are not based on independent samples. For example,  $p_1$  may represent the proportion of individuals who give a particular response prior to a provided experience and  $p_2$  the proportion of the *same* individuals giving the response after the interpolated experience; or  $p_1$  might be the proportion passing one test item and  $p_2$  the proportion of the same group passing a second item, and we wonder whether items differ significantly as to difficulty; or we may wish to test the significance of the difference

between the responses of a group to two opinion questions; or we may need to evaluate the difference between two proportions based on paired or matched individuals.

For such situations as the foregoing, neither formula (1) nor (2) is applicable. To be applicable, a subtractive correlational term should be included. The needed correlation is that between the sampling variations of the two proportions. Since it can readily be shown that proportions are means, and since it is known that the correlation between sample means equals that between variates, it follows that the needed correlation is that between first and second responses, or between performances on the two items being compared for difficulty, or between responses to two opinion questions, or between the responses of matched cases. The correlation "scatter plot" for these situations involves a two-by-two table, for which one may apparently choose either the contingency coefficient or tetrachoric  $r$  or the four-fold point  $r$  as the measure of correlation. Which type of coefficient do we need for the sampling variance of the difference between correlated proportions?

If the answer to this question is to be found in the statistical literature, the writer has failed to locate it. If the answer were obvious, one would not expect the near universal failure of researchers to include the needed correlational term in those situations where it must certainly be taken into account in testing significance. We have found only one text\* which considers the correlational term, and the authors of this text, after saying that "it is seldom that we know the correlation factor when dealing with proportions," go on to state without proof that "it would need to be computed by the tetrachoric method." We shall presently see that this idea is incorrect.

The desired sampling variance formula would be of the form

$$\sigma_d^2 = \sigma_{p_1}^2 + \sigma_{p_2}^2 - 2r \frac{\sigma_{p_1 p_2}}{p_1 p_2} \frac{\sigma_{p_1}}{p_1} \frac{\sigma_{p_2}}{p_2}, \quad (3)$$

in which  $r$  is not as yet specified. Suppose a fourfold table of frequencies, and the corresponding table of proportions, for a first (I) and a second (II) set of responses from the same individuals (see Table 1). The difference between the two marginal values,  $p_2 - p_1$ , is the difference for which we seek a standard error. For the sake of exposition, let us assume that this difference represents a change in responses. Then the  $B$  and  $C$  individuals are those who have not changed, while  $A$  represents those who moved from yes to no, and  $D$  those who changed from no to yes. The lower-case letters represent the corresponding proportions. Let us score those who changed to

\* Peters, C. C. and Van Voorhis, W. R. Statistical procedures and their mathematical bases. New York: McGraw-Hill, 1940. P. 183.

TABLE 1  
Fourfold Table of Original and Proportional Frequencies

(I)	(II)			(I)	(II)		
	No	Yes			No	Yes	
	Yes	No			Yes	No	
	A	B	A+B		a	b	p <sub>1</sub>
	C	D	C+D		c	d	q <sub>1</sub>
	A+C	B+D	N		q <sub>2</sub>	p <sub>2</sub>	1.00

yes as +1, those who changed to no as -1, and those who changed not at all as 0. This leads to the frequency distributions for changes shown in Table 2, which includes easily derived expressions for the mean and variance of the distributions of changes.

TABLE 2  
Distributions for Changes in Terms of Original and Proportional Frequencies

Original		Proportional	
X	f	X	f
+1	D	+1	d
0	B+C	0	b+c
-1	A	-1	a
	N		1.00

$$\text{Mean} = (D-A)/N = d-a = p_2 - p_1$$

$$S.D.^2 = \frac{1}{N^2} [N(D+A) - (D-A)^2] = (d+a) - (d-a)^2$$

Obviously, the mean change equals  $(D-A)/N = (d-a) = p_2 - p_1$ , or the mean of the changes (or differences) equals the difference between the two proportions. The sampling error of  $p_2 - p_1$  must be the same as that for  $M$ , its equivalent, and the sampling variance of  $M$  is easily obtained by dividing the distribution variance by  $N$ . Thus, in terms of proportional frequencies, we have

$$\sigma_M^2 = \sigma_{p_2-p_1}^2 = \frac{1}{N} [(d+a) - (d-a)^2]. \quad (4)$$

Now by easy, but somewhat cumbersome algebra which need not be reproduced here, it can be shown that (4) reduces to (3) *providing* the  $r$  of (3) is the fourfold point correlation coefficient. It will be noted that (4) requires less computation than (3) even when  $r$  has been calculated for its own sake.

It will be seen that (3), hence by implication that (4), involves the use of two separate estimates of the unknown population proportions, and therefore both are analogous to (1). If we wish a formula analogous to the preferred (2), that is, one involving an estimate based on the combined sets of responses or a  $p$  based on the total as a better estimate, we note that such a  $p$  would, as in the case of (2), imply that the population values for the two proportions are equal. This would be consistent with the null hypothesis that  $p_2 - p_1 = 0$  except for random sampling fluctuations. The proper  $p$  to replace  $p_2$  and  $p_1$  in (3) becomes the simple average of  $p_2$  and  $p_1$  since for the given situation  $N_2 = N_1 = N$ . Now the null hypothesis that  $p_2 = p_1$  also implies that the equivalent,  $d - a$ , must equal zero except for sampling errors; hence formula (4) becomes

$$\sigma^2_{p_2-p_1} = \frac{1}{N} (d + a), \quad (5)$$

which is algebraically identical to (3) with  $p_2$  and  $p_1$  replaced by  $p$ . Since (2) is associated exactly with  $\chi^2$ , it is of interest to point out the connection of (5) with  $\chi^2$ . Table 3 contains the setup for  $\chi^2$  for the original observed ( $O$ ) frequency distribution of Table 2. Since the net

TABLE 3  
Schema for  $\chi^2$  Test of Changes

$X$	$O$	$E$	$(O-E)^2/E$
+1	$D$	$(D+A)/2$	$\frac{(D-A)^2}{4} \cdot \frac{2}{D+A} = \frac{1}{2} \frac{(D-A)^2}{D+A}$
0	$B+C$		
-1	$A$	$(D+A)/2$	$\frac{(A-D)^2}{4} \cdot \frac{2}{D+A} = \frac{1}{2} \frac{(D-A)^2}{D+A}$

change (or difference) between the two sets of responses depends upon the  $D$  and  $A$  individuals, the expected frequencies ( $E$ ) are written on the assumption that the  $D + A$  cases would be divided evenly, on the basis of the null hypothesis, between plus and minus changes. The value of  $\chi^2$  becomes the sum of the right-hand terms, or

$$\chi^2 = \frac{(D-A)^2}{D+A}. \quad (6)$$

When we write the square of the critical ratio with (5) as the sampling variance, we have

$$(CR)^2 = \frac{(d-a)^2}{\frac{d+a}{N}}.$$

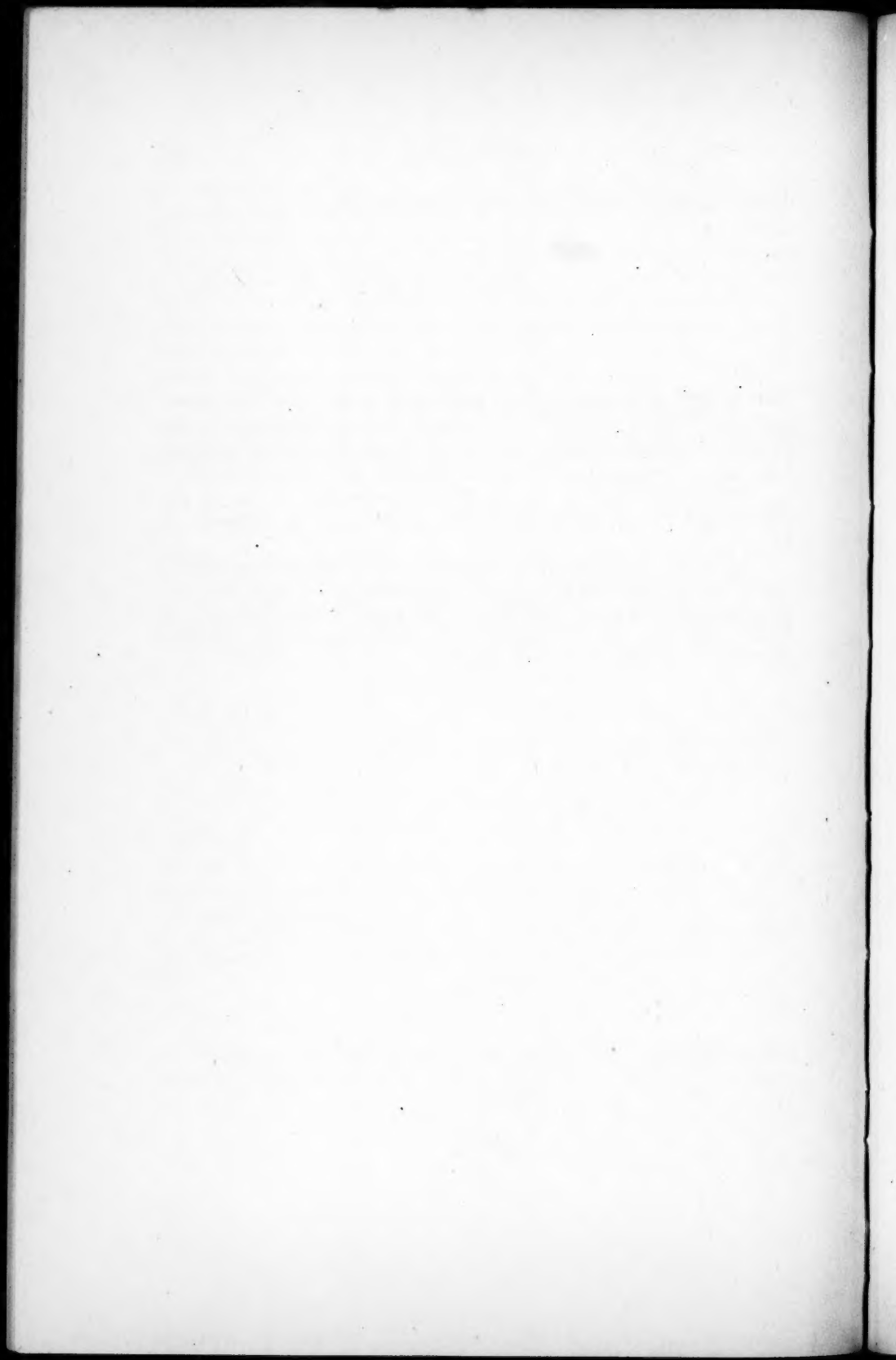
Replacing proportions by frequencies, this becomes

$$\frac{\left(\frac{D}{N} - \frac{A}{N}\right)^2}{\frac{\frac{D}{N} + \frac{A}{N}}{N}} = \frac{(D - A)^2}{D + A}.$$

Thus  $(CR)^2$ , using formula (5), is identical to the  $\chi^2$  of (6). Obviously,  $CR = (D - A)/\sqrt{D + A}$ , and therefore the significance of the difference between  $p_2$  and  $p_1$  can readily be computed from the fourfold table of frequencies or from the fourfold table of proportions.

The use of either (5) or (6) is to be preferred over (3) or (4) just as, and for the same reason that, formula (2) is preferable to (1).

Although the foregoing development was in terms of changes in responses, the resulting formulas are applicable to the other three situations mentioned at the beginning of this note. The limitation on chi square regarding size of expected frequencies suggests that formulas (6) and (5) should not be used unless  $A + D$  is 10 or greater.



# REVIEWS

HENRY B. MANN. Nonparametric tests against trend. *Econometrica*, 1945, 13, 245-259.

This paper provides a statistical test of the hypothesis that a sequence of measures does not show a downward trend. The test against upward trend can be made in a completely analogous manner. The test is nonparametric in that no assumption is made about the distribution of the measurements. The test is based upon the rank order of  $n$  measurements,  $X_1, X_2, \dots, X_n$ , considered as a member of the class of permutations of the first  $n$  integers. Each of the permutations is considered equally probable.

The first or  $T$  test is based upon the number of inversions occurring among the  $n$  measurements in the sample. An inversion is a situation where  $X_i < X_j$ , and  $i < j$ .  $T$ , the number of such inversions, is the sample statistic. The probability of obtaining a permutation of the first  $n$  integers such that  $T < \bar{T}$  is tabled for  $\bar{T} = 0, 1, \dots, 21$ , and  $n = 3, 4, \dots, 10$ . For  $n > 10$ , the probability may be estimated from the normal curve.

If the hypothesis of randomness is true, the probability of any  $X_i$  being less than any  $X_k$  when  $i < k$  is .50. The  $T$  test is shown to be consistent with respect to any of a wide class of alternative hypotheses when this probability is different from .50, less than .50 on the average, and when all measurements are independent. By consistent it is meant that the probability of rejecting the hypothesis of randomness when some other hypothesis among the prescribed alternatives is true, approaches certainty as the size of the sample increases without limit.

A second test of the hypothesis is also presented. It is more powerful than the  $T$  test if the probability that  $X_i > X_j$  increases rapidly with  $j-i$ . In the sample,  $X_1, X_2, \dots, X_n$ , it may always be true that  $X_i > X_j$ , if  $i < j$  and the difference in ranks is  $K$  or more. The smallest value of  $K$  for which that statement is true is the sample statistic. The probability of various values of  $K$  for any size sample under the hypothesis of randomness is tabled in the article.

FELS INSTITUTE FOR RESEARCH IN HUMAN DEVELOPMENT

ALFRED L. BALDWIN

GERHARD TINTNER. Multiple regression for systems of equations. *Econometrica*, 1946, 14, 5-36.

Tintner in this publication faces the general problem of estimating the number of independent relations among a group of economic variables, and further to estimate the coefficients in the following set of equations which describe these relationships.

$$k_{v0} + \sum_{j=1}^p k_{vj} M_{jt} = 0 \quad v=1, 2, \dots, R \quad t=1, 2, \dots, N.$$

The  $R$  such equations represent the  $R$  independent linear relationships among

the variables. Each such equation, or set of equations, will relate the  $p$  variables, indicated by the  $j$  subscript, at the various times  $t = 1, 2, \dots$ . To estimate the number,  $R$ , of independent relationships and the values of  $k_{vj}$ , the sample contains values  $X_{jt}$ , of the  $p$  variables at  $N$  times. The sample statistics needed for the statistical solution are the means and variances of each variable over the time series, and the covariance of every pair of variables over the time series. Since time series are being used, the estimation of the variance must include only the random portion of the temporal data. Tintner's own variate difference method may be used to obtain estimates of the variance of each variable, although its use is not integral to the discussion.

The statistical solution Tintner proposes closely resembles the factor analysis of a matrix of variances and covariances by the method of principal components. What is required is first a solution of the determinantal equation  $|a_{ij} - \lambda V_i| = 0$  and then a substitution of the roots of this equation to obtain values of  $k_{vj}$ , the structural coefficients.  $V_i$  is the variance of each variable as determined by the method of variate differences or some similar method.

Using a method developed by Hsü, Tintner, determines the rank of the matrix. He calculates all the principal components of the matrix by Hotelling's method, and then uses the following test function:

$$\Lambda_r = (N-1) (\lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_r),$$

where  $\lambda_1$  is the smallest root,  $\lambda_2$  the next smallest, etc. This function is distributed as  $\chi^2$  with  $(N-1-p+r)$  degrees of freedom if the rank of the matrix is  $r$ . By calculating successively  $\lambda_1, \lambda_2, \dots$  until the hypothesis rank  $= r$  is rejected, the rank is estimated at  $R$ , the largest value of  $r$  for which the hypothesis is acceptable.

By substituting the  $R$  smallest roots of  $|a_{ij} - \lambda V_i| = 0$  into the equation, and by further subjecting the coefficients to the conditions:

1.  $\sum_{i=1}^p k_{vi} k_{wi} V_i = \delta_{vw}$ , where  $\delta_{vw}$  is Kronecker delta, and
2.  $\sum_{t=1}^N (\sum_{j=1}^p k_{vj} x_{jt}) (\sum_{j=1}^p k_{wj} x_{jt}) = 0$ ,

$R$  linear functions are obtained describing the  $R$  linear relationships among the variables.

At this point Tintner has mathematical expressions of the interrelations, but they are not very meaningful in terms of economics. To identify each relationship, he recalculates it under the assumption that some one or more of the coefficients in that relationship are equal to zero. This assumption is made on the basis of economic theory. For example, he assumes that in a "demand" curve for agricultural products, the prices paid by farmers has no weight, i.e., the coefficient of that term in the equation is equal to zero.

Making these assumptions generally reduces the fit of the curves to the data, but the reduction in fit can be tested. The ratio,  $\Lambda_r/\Lambda'_r$ , under different assumptions is approximately distributed as  $F$  if the fit under the different assumptions is equally good. Not only can the general fit be evaluated statistically, but each coefficient can be tested individually. The method is illustrated by a problem including five variables describing agricultural data.

The method is suitable only for large samples at present. If the number of

variables were at all large, the computation would be excessively laborious, because a complete factor analysis into principal components is necessary and a calculation of the inverse of the matrix of variances and covariances is required. On the other hand the existence of a statistical test for rank and for the coefficients is highly desirable and might make the method worth the additional computation—additional, that, is to that involved in some other method of factor analysis. While the  $k_{vj}$  are not factor loadings, some of the methods used could be adapted to the usual factor-analysis problem.

ALFRED L. BALDWIN

FELS INSTITUTE FOR RESEARCH IN HUMAN DEVELOPMENT



MEMBERS OF THE PSYCHOMETRIC SOCIETY\*

- William Michael Adams, c/o Evelyn Brandt, 5502 - 15 St., N.E., Seattle, Washington
- J. C. Adcock, Victoria University College, Wellington, New Zealand.
- Dorothy C. Adkins, 2032 Belmont Rd., N.W., Washington 9, D.C.
- Robert B. Ammons, School of Education, Syracuse University, Syracuse, New York
- Wade S. Amstutz, 8 Fritchie Pl., Dayton 10, Ohio
- Gordon V. Anderson, Dept. of Psychology, Northwestern University, Evanston, Illinois
- T. Gaylord Andrews, University of Chicago, Chicago, Illinois
- Chester Banachowski, 121 - 31st Ave., San Mateo, California
- A. S. Barr, University of Wisconsin, Madison 6, Wisconsin
- Brent Baxter, Owens-Corning Fiberglass Co., Nicholas Blvd., Toledo 1, Ohio
- Harold P. Bechtoldt, 1150½ East 61 St., Chicago 37, Illinois
- Ralph C. Bedell, 322 Teachers College, University of Nebraska, Lincoln, Nebraska
- Roger M. Bellows, Wayne University, Detroit, Michigan
- George K. Bennett, Psychological Corp., 522 - 5th Ave., New York 18, New York
- Corlin O. Beum, 2000-A 38 St., S.E., Washington 20, D.C.
- W. V. Bingham, 1661 Crescent Pl., Washington 9, D.C.
- Ruth Bishop, 239 West 20th St., New York 11, New York
- Reign H. Bittner, 3133 Copland Blvd., Toledo 9, Ohio
- Paul Blommers, State University of Iowa, Iowa City, Iowa
- Frances Blumenthal, 6 Landscape Ave., Yonkers 5, New York
- Edward S. Bordin, Student Counseling Center, Washington State College, Pullman, Washington
- Henry Borow, The General College, University of Minnesota, Minneapolis 14, Minnesota
- Robert E. Breden, Personnel Research Section, Procter and Gamble Co., Sixth and Main Sts., Cincinnati, Ohio
- Claude F. Bridges, 6 Livingston Ave., Yonkers 5, New York
- Hubert E. Brogden, Pers. Res. & Proc. Br., AGO, War Dept., Pentagon, Washington 25, D.C.
- Cecil R. Brolyer, 3121 - 38th St., N.W., Washington, D.C.
- Theodore H. Brown, Harvard Business School, Boston 63, Massachusetts
- Mrs. Virginia M. Brown, 5800 Blackstone Ave., Apt. 3, Chicago 37, Illinois
- Joseph G. Bryan, 97 Green St., Melrose, Massachusetts
- Hermann F. Buegel, University of North Dakota, University Station, Grand Forks, North Dakota
- Oscar K. Buros, 220 Montgomery St., Highland Park, New Jersey
- Harold E. Burt, Dept. of Psychology, Ohio State University, Columbus 10, Ohio

\* An asterisk preceding a name indicates student membership.

- Hobart Bushey, Math. Dept., Hunter College, Lexington Ave. & 68th St., New York 21, New York
- John B. Carroll, 3325 20th Rd., N., Arlington, Virginia
- Harold D. Carter, 10 East Parnassus St., Berkeley 8, California
- Launor F. Carter, Dept. of Psychology, University of Rochester, Rochester 3, New York
- Teobaldo Casanova, University of Puerto Rico, Rio Piedras, Puerto Rico
- Jozef B. Cohen, Dept. of Psychology, Cornell University, Ithaca, New York
- Herbert S. Conrad, 67 Jefferson Rd., Princeton, New Jersey
- Clyde H. Coombs, Univ. of Michigan, 1027 E. Huron St., Ann Arbor, Michigan
- Herman A. Copeland, 543 Overlook Rd., Philadelphia 28, Pennsylvania
- Gertrude M. Cox, North Carolina State College of Agriculture and Engineering, Raleigh, North Carolina
- H. M. Cox, University of Nebraska, Lincoln 8, Nebraska
- Clarke W. Crannell, Dept. of Psychology, Miami University, Oxford, Ohio
- William J. E. Crissey, 51-01 39th Ave., Sunnyside, Long Island City, New York
- Lee J. Cronbach, Dept. of Education, University of Chicago, Chicago 37, Illinois
- Dorothea J. Crook, 494 E. 74th St., New York 21, New York
- Elmer K. Culler, Laboratory of Psychology, River Campus, University of Rochester, Rochester 3, New York
- Edward E. Cureton, Pers. Res. & Proc. Br., AGO, War Dept., Pentagon, Washington 25, D.C.
- Manuel Cynamon, 10 Ave. P, Brooklyn 4, New York
- John T. Dailey, Psychol. Res. and Examining Unit, Sqn. E, Indoctrination Division, Air Training Command, San Antonio, Texas
- Karl M. Dallenbach, 102 Irving Pl., Ithaca, New York
- Frederick B. Davis, 616 Poplar Dr., Falls Church, Virginia
- Hugh M. Davison, 702 Wade Ave., Rockville, Maryland
- Walter L. Deemer, Jr., School of Aviation Medicine, Randolph Field, Texas
- \*James W. Degan, 6007 S. Drexel Ave., Chicago 37, Illinois
- \*L. C. Deschene, University of Houston, Houston 4, Texas
- Jerome E. Doppelt, 220 East 204 St., New York 58, New York
- \*Charlotte G. Dragstedt, 5200 Greenwood Ave., Chicago 15, Illinois
- Philip H. DuBois, Dept. of Psychology, Washington University, St. Louis 5, Missouri
- Jack W. Dunlap, Psychological Corporation, 522 - 5th Ave., New York 18, New York
- Paul S. Dwyer, 640 Oxford Rd., Ann Arbor, Michigan
- Henry Dyer, 458 Pleasant St., Belmont, Massachusetts
- Harold A. Edgerton, Dept. of Psychology, Ohio State University, Columbus 10, Ohio
- Wm. G. Emmet, Moray House, Edinburgh 8, Scotland
- Max D. Engelhart, 5427 Greenwood Ave., Chicago 15, Illinois

- Mrs. Katherine W. Estes, Dept. of Psychology, Indiana University, Bloomington, Indiana
- Leonard W. Ferguson, Field Management Division, Metropolitan Life Insurance Co., 1 Madison Ave., New York 10, New York
- Samuel W. Fernberger, College Hall, University of Pennsylvania, Philadelphia 4, Pennsylvania
- George H. Fettus, 5701 Pontiac Rd., Berwyn Heights, Maryland
- Warren G. Findley, Educational Advisory Staff, Air University, Maxwell Field, Alabama
- P. F. Finner, Florida State College for Women, Tallahassee, Florida
- John C. Flanagan, Dept. of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania
- \*Benjamin Fruchter, 4318 Melburne Dr., Los Angeles 33, California
- Edmund F. Fuchs, Pers. Res. & Proc. Br., AGO, War Dept., Pentagon, Washington 25, D.C.
- N. L. Gage, Division of Educational Reference, Purdue University, Lafayette, Indiana
- Henry E. Garrett, Dept. of Psychology, Columbia University, New York 27, New York
- \*Frederick Gehlman, 7427 - 3K South Shore, Chicago 49, Illinois
- Charles P. Gershenson, Box 130, Teachers College, Columbia University, New York 27, New York
- Jerome Gibson, 336 Richmond Ave., Dayton 6, Ohio
- John G. Gleason, 706 North 18th St., Lafayette, Indiana
- \*Alfred F. Glixman, Dept. of Psychology, University of California, Berkeley, California
- David A. Grant, Dept. of Psychology, University of Wisconsin, Madison 6, Wisconsin
- J. Douglas Grant, 609 Homer Ave., Palo Alto, California
- Clyde H. Graves, 23 Oak Ave., Hempstead, Long Island, New York
- Thomas N. Greville, 1714 - 34th St., N.W., Washington 7, D.C.
- William W. Grings, Dept. of Psychology, University of Denver, Denver, Colorado
- Harold Guetzkow, Psychology Dept., University of Michigan, Ann Arbor, Michigan
- J. P. Guilford, 509 North Rexford Dr., Beverly Hills, California.
- Harold O. Gulliksen, Box 592, Princeton, New Jersey
- Edwin R. Guthrie, 3914 Belvoir Pl., Seattle 5, Washington
- Louis Guttman, Dept. of Sociology, Cornell University, Ithaca, New York
- Harry H. Harman, 4111 Maryland Dr., Washington, D.C.
- T. W. Harrell, University of Illinois, Urbana, Illinois
- Charles M. Harsh, Dept. of Psychology, University of Nebraska, Lincoln 8, Nebraska
- L. D. Hartson, Oberlin College, Oberlin, Ohio
- Jack H. Hazelhurst, 343 South Dearborn St., Chicago 4, Illinois
- E. R. Henry, 428 Demarest Ave., Oradell, New Jersey

- Conwell Higgins, 406 Hudson Ave., Albany 3, New York  
Elmer D. Hinckley, Box 2007, University Station, Gainesville, Florida  
Karl J. Holzinger, Dept. of Education, University of Chicago, Chicago 37, Illinois  
A. Paul Horst, Procter and Gamble Co., 6th and Main Sts., Cincinnati 1, Ohio  
Alston S. Householder, Monsanto Chemical Co., Clinton Laboratories, Box 1991,  
Knoxville, Tennessee  
Doncaster G. Humm, Box 1433, Del Valle Station, Los Angeles 15, California  
Rheem F. Jarrett, 1800 San Antonio, Berkeley 7, California  
Nathan Jaspen, 1355 Queen St., N.E., Washington 2, D.C.  
Palmer O. Johnson, Dept. of Education, University of Minnesota, Minneapolis 14,  
Minnesota  
Clifford E. Jurgensen, Personnel Director, Minneapolis Gas Light Co., Minneapolis,  
Minnesota  
John E. Karlin, Bell Telephone Laboratories, Murray Hill, New Jersey  
Robert J. Keller, 1810 Rogers Ave., San Antonio 8, Texas  
T. L. Kelley, Walker House, 40 Quincy St., Cambridge 38, Massachusetts  
Mary Josephine Kientzle, Box 804, College Station, Pullman, Washington  
Joseph King, Science Research Association, 228 S. Wabash Ave., Chicago 4, Illinois  
Helen Lois Koch, 1374 East 57th St., Chicago 37, Illinois  
Joseph F. Kubis, Fordham University, Graduate School, New York 58, New York  
G. Frederic Kuder, 917 - 15th St., N.W., Washington 5, D.C.  
Albert K. Kurtz, Box 282, East Lansing, Michigan  
George M. Kuznets, Giannini Foundation, University of California, Berkeley 4,  
California  
John I. Lacey, Samuel S. Fels Research Institute, Yellow Springs, Ohio  
Oliver L. Lacey, University of Alabama, University, Alabama  
Herbert D. Landahl, 5822 Drexel Blvd., Chicago 37, Illinois  
Charles R. Langmuir, 125 Spencer Pl., Mamaroneck, New York  
Paul F. Lazarsfeld, Office of Radio Research, 15 Amsterdam Ave., New York 23,  
New York  
D. Welty Lefever, University of Southern California, University Park, Los Angeles 7, California  
Theo. F. Lentz, Jr., Character Research Institute, Washington University, St. Louis 5, Missouri  
Hollis M. Leverett, 21 Hampden St., Wellesley, Massachusetts  
Ernest M. Ligon, Union College, Schenectady 8, New York  
E. F. Lindquist, State University of Iowa, Iowa City, Iowa  
Irving D. Lorge, 525 West 120th St., New York 27, New York  
Maurice Lorr, 7703 Eastern Ave., N.W., Washington 12, D.C.  
Ardie Lubin, 6136 Kenwood, Chicago 37, Illinois  
Bertha M. Luckey, Psychology Clinic, 1380 East 6th St., Cleveland 14, Ohio  
Brother Luke, Mont-de-la-Salle, Laval Rapids, P.Q., Canada  
George A. Lundberg, Dept. of Sociology, University of Washington, Seattle, Washington

Walter A. Lurie, Natl. Comm. Rel. Adv. Counc., 295 Madison Ave., New York 17, New York

Benjamin Malzberg, Dept. of Mental Hygiene, State Office Building, Albany 1, New York

Octavio A. L. Martins, Rua Figueredo Magalhaes 27, Apt. 702, Rio de Janeiro, Brazil

William G. Matheny, University of Maryland, College Park, Maryland

Luzelle Denton Mays, c/o Luther Mays, Dept. of Psychology, Syracuse University, Syracuse, New York

\*William A. McClelland, Dept. of Psychology, University of Minnesota, Minneapolis 14, Minnesota

C. H. McCloy, Field House, Iowa City, Iowa

Walter J. McNamara, Dept. of Education, International Business Machines Corp., Endicott, New York

Quinn McNemar, Dept. of Psychology, Stanford University, California

\*Francis F. Medland, 506 N. Highland St., Arlington, Virginia

Ralph Carl Meister, 15 North Root St., Aurora, Illinois

Paul Mendenhall, Boy Scouts of America, 2 Park Ave., New York 16, New York

Lorenz A. Meyer, 5031 First St., N.W., Washington, D.C.

Thomas Verner Moore, Benedictine Foundation, Washington 17, D.C.

Charles I. Mosier, 2032 Belmont Rd., N.W., Washington 9, D.C.

Carl Murchison, The Journal Press, 2 Commercial St., Provincetown, Massachusetts

Paul M. Neurath, 510 West 124th St., New York 27, New York

Wilson L. Newman, 5548 South Blackstone Ave., Chicago 37, Illinois

James A. Norton, Jr. (Address not known)

Ewald R. Nyquist, 550 Winwood Pl., Pelham Manor, Pelham, New York

Willard C. Olson, University Elementary School, University of Michigan, Ann Arbor, Michigan

Katharine Pease, Central Dr., Briarcliff Manor, New York

Donald A. Peterson, College Entrance Examination Board, 2½ Chambers St., Princeton, New Jersey

B. Roger Philip, 220 Pine Ave., West, Montreal, Canada

H. Pieron, College de France, Place Marcelin Berthelot, Paris, France

Lynette Birdsall Plumalee, Box 592, Princeton, New Jersey

John H. Rappaport, U. S. Rubber Co., 6600 East Jefferson Ave., Detroit 32, Michigan

Wilhelm Reitz, 960 Atkinson St., Detroit 2, Michigan

Martin L. Reymert, Mooseheart Lab. for Child Research, Mooseheart, Illinois

Marion W. Richardson, 56 Beaver St., New York 4, New York

\*Arthur Riopelle, Dept. of Psychology, University of Wisconsin, Madison, Wisconsin

- John Rohrs, 15 Tamarack Rd., Staten Island 4, New York  
Joseph R. Royce, 1147 East 61st St., Chicago, Illinois  
Phillip J. Rulon, Walker House, 40 Quincy St., Cambridge 38, Massachusetts  
T. A. Ryan, Dept. of Psychology, Cornell University, Ithaca, New York  
Douglas E. Scates, College Station, Durham, North Carolina  
Willis C. Schaefer, Dept. of Psychology, University of Maryland, College Park, Maryland  
George E. Schlessner, 37 Hamilton St., Hamilton, New York  
William B. Schrader, Dept. of Psychology, University of Tennessee, Knoxville, Tennessee  
Douglas Schultz, College Entrance Examination Board, 2½ Chambers St., Princeton, New Jersey  
Saul B. Sells, 306 Abisco Ave., San Antonio 2, Texas  
Robert B. Selover, Prudential Insurance Co. of America, Newark 1, New Jersey  
Marion F. Shaycroft, 16 Prospect Park, S.W., Brooklyn 15, New York  
Joseph Russell Sherlock, 2685 Briggs Ave., New York 58, New York  
Nathan Shock, Baltimore City Hospital, Unit on Gerontology, Baltimore 24, Maryland  
Paul S. Siegel, Dept. of Psychology, University of Alabama, University, Alabama  
George B. Simon, Office of the Surgeon, Hq. Air Training Command, Barksdale Fd., Louisiana  
Ralph J. Slaterry, Eldridge, California  
Anthony J. Smith, Dept. of Psychology, University of Kansas, Lawrence, Kansas  
Max Smith, 453 East 96th St., Brooklyn 12, New York  
\*Allen J. Sprow, Institute of Human Relations, 333 Cedar St., New Haven, Connecticut  
John M. Stalnaker, 111 Martin Ave., Palo Alto, California  
Daniel Starch, 1850 Graybar Building, 420 Lexington Ave., New York 17, New York  
Arthur I. Sternhell, Field Management Division, Metropolitan Life Insurance Co., 1 Madison Ave., New York 10, New York  
Naomi Stewart, 1925 Glenwood Rd., Brooklyn 30, New York  
George D. Stoddard, 711 W. Florida Ave., Urbana, Ill.  
Samuel A. Stouffer, 38 Hastings Rd., Belmont 78, Massachusetts  
Edward G. Stoy, Rt. 1, Box 42, San Louis Obispo, California  
Edward K. Strong, Jr., 762 Mirada, Stanford University, California  
Reginald L. Swann, Teachers College of Connecticut, New Britain, Connecticut  
Frances Swineford, Dept. of Education, University of Chicago, Chicago 37, Illinois  
Calvin W. Taylor, 1813 South 15th St., East, Salt Lake City, Utah  
E. K. Taylor, Pers. Res. & Proc. Br., AGO, War Dept., Pentagon, Washington 25, D. C.  
Harold C. Taylor, Director, W. E. Upjohn Institute for Community Research, American National Bank Building, Kalamazoo 4, Michigan  
L. Grant Tennies (Address not known)

- Lewis M. Terman, 761 Dolores St., Stanford University, California  
Godfrey H. Thomson, Moray House, University of Edinburgh, Edinburgh, Scotland  
E. L. Thorndike, 501 West 120th St., New York 27, New York  
Robert L. Thorndike, Gull Rock Rd., Madison, Connecticut  
L. L. Thurstone, University of Chicago, Chicago 37, Illinois  
Thelma G. Thurstone, 5642 Kimbark Ave., Chicago 37, Illinois  
David V. Tiedeman, Walker House, 40 Quincy St., Cambridge 38, Massachusetts  
Herbert A. Toops, 1430 Cambridge Blvd., Columbus 8, Ohio.  
Arthur E. Traxler, 437 West 59th St., New York 19, New York  
Ledyard R. Tucker, 50½ Maple St., Princeton, New Jersey  
Ralph W. Tyler, Dept. of Education, University of Chicago, Chicago 37, Illinois  
Howard F. Uphoff, 6400 Elliott Pl., Hyattsville, Maryland  
C. C. Upshall, 78 Oneta Rd., Rochester 5, New York  
Marvin J. Van Wagenen, University of Minnesota, Minneapolis 14, Minnesota  
Charles L. Vaughn, Psych. Corp., 522 5th Ave., New York 18, New York  
David F. Votaw, Dept. of Education, Southwest Texas State Teacher's College, San Marcos, Texas  
Helen M. Walker, Teachers College, Columbia University, New York 27, New York  
M. J. Wantman, Catherine Strong Hall, University of Rochester, Rochester 3, New York  
Neil D. Warren, 3827 Olympiad Dr., Los Angeles 43, California  
John G. Watkins, Dept. of Psychology, Washington State College, Pullman, Washington  
R. R. G. Watt, 1188 West 28th St., Los Angeles, California  
David Wechsler, 145 East 92nd St., New York 28, New York  
Alexander G. Wesman, 20 Seaman Ave., New York 34, New York  
Robert J. Wherry, Pers. Res. & Proc. Br., AGO, War Dept., Pentagon, Washington 25, D. C.  
Martin M. White, 173 Cherokee Park, Lexington 1, Kentucky  
S. S. Wilks, Dept. of Mathematics, Princeton University, Princeton, New Jersey  
J. C. Williamson, Cheyney, Pennsylvania  
Louise R. Witmer, 3130 Wisconsin Ave., N.W., Washington 16, D.C.  
Dael L. Wolfle, 2801 Rittenhouse St., N.W., Washington 15, D. C.  
Ben D. Wood, Columbia University, New York 27, New York  
Herbert Woodrow, Dept. of Psychology, University of Illinois, Urbana, Illinois  
Ella Woodyard, Wendover, Leslie County, Kentucky  
Ruth E. Wright, 5479 University Ave., Chicago 15, Illinois  
J. Wayne Wrightstone, 21 Hickory Rd., Summit, New Jersey  
Theodore O. Yntema, 5729 Blackstone Ave., Jackson Park Station, Chicago 37, Illinois  
Joseph Zubin, 722 W. 168th St., New York, New York



